Pangenomic Landscapes Shape Performances of a Synthetic Genetic Circuit Across *Stutzerimonas* Species

Dennis Tin Chat Chan[1], Hans C. Bernstein[1,2]*

[1]Faculty of Biosciences, Fisheries and Economics, UiT - The Arctic University of Norway, 9019, Tromsø, Norway

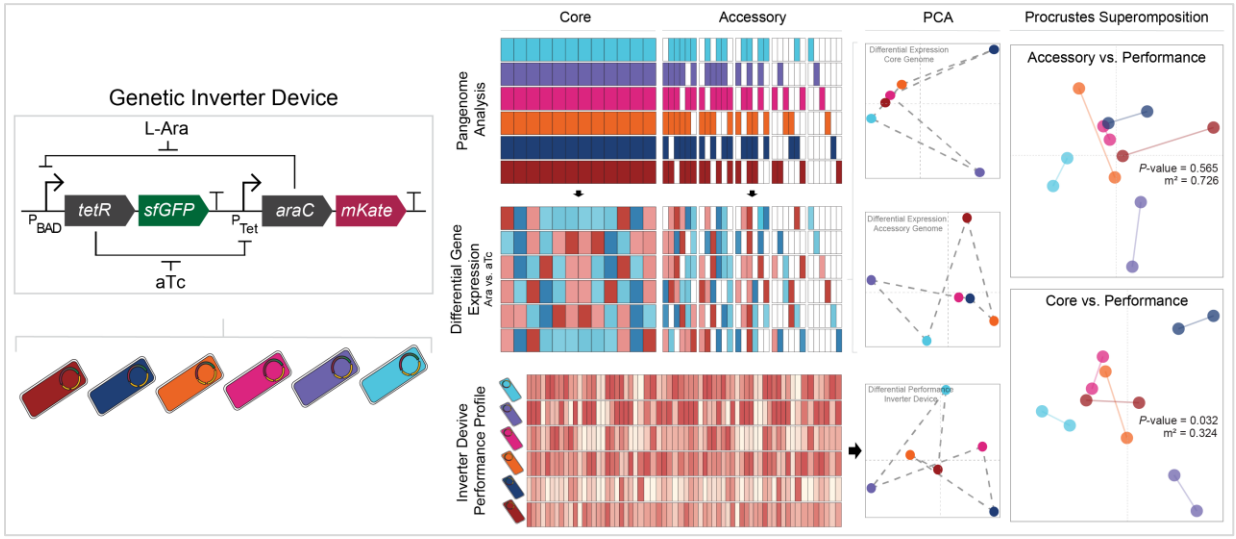
[2]The Arctic Centre for Sustainable Energy, UiT - The Arctic University of Norway, 9019, Tromsø, Norway

* Correspondence: Hans. C. Bernstein (hans.c.bernstein@uit.no)

Keywords:

Synthetic Biology, Genetic Inverter, Transcriptomics, Chassis-Effect, Biodesign, Procrustean Superimposition, Context Dependence, Non-Model Organism

GRAPHICAL ABSTRACT



ABSTRACT

Engineering identical genetic circuits into different species typically results in large differences in performance due to the unique cellular environmental context of each host, a phenomenon known as the "chassis-effect". A better understanding of how genomic and physiological contexts underpin the chassis-effect will greatly improve biodesign strategies across diverse microorganisms. Here, we combined a pangenomics-based gene expression analysis with quantitative measurements of performance from an engineered genetic inverter device to uncover how genome structure and function relates to the observed chassis-effect across six closely related *Stutzerimonas* hosts. Our results reveal that genome architecture underpins divergent responses between our chosen non-model bacterial hosts to engineered genetic circuits. Specifically, differential expression of the core genome, gene clusters shared between all hosts, were found to be the main source of significant concordance to the observed genetic device performance, whereas specialty genes from respective accessory genomes were not significant. A data-driven investigation revealed that genes involved in denitrification and components of trans-membrane transporter proteins were among the most differentially expressed gene clusters in response to the genetic device. Our results show the chassis-effect can be traced along differences among genome-encoded functions that are mostly conserved and that these differences create a unique biodesign space among closely related species.

IMPORTANCE

Contemporary synthetic biology endeavors often default to a handful of model organisms to host their engineered systems. Model organisms such as *Escherichia coli* serve as attractive hosts due to their tractability but do not necessarily provide the ideal environment to optimize performance, and as more novel microbes are domesticated for use at biotechnology platforms, synthetic biologists are urged to explore the chassis-design space in order to optimize their systems and deliver on the promises of synthetic biology. Therefore, the consequences of the chassis-effect will only become more relevant as the field of biodesign grows. In our work, we demonstrate that the performance of a genetic device is highly dependent on the host environment it operates within, promoting the notion that the chassis can be considered a design variable to tune circuit function. Importantly, our results unveil that the chassis-effect can be traced along similarities in genome architecture, specifically the shared core genome. Our study advocates for the exploration of the chassis-design space and is a step forward to empowering synthetic biologists with knowledge for more efficient exploration of the chassis-design space to enable the next generation of broad-host-range synthetic biology.

INTRODUCTION

As synthetic biologists continue to explore the design space of engineered genetic circuits, we are presented with a complex landscape where functional fidelity depends on host physiology, environment, and genetic tractability[1,2]. This has given rise to a subdiscipline of broad-host-range synthetic biology, which aims to create versatile genetic systems that can operate across diverse organisms, and is particularly beneficial towards biotechnologies that capitalize on microbial diversity[3,4]. Despite the plethora of modular genetic parts available, we are still faced with the challenge that identical genetic devices exhibit consequentially different performance depending on the host context the device operates in, a process termed the "chassis-effect"[5,6] or "context dependence"[7,8]. The chassis-effect hinders the predictability of circuit function based on part composition alone and can cause any circuit optimization typically done in a model organism (e.g., *Escherichia coli*) to be rendered null once introduced into a different host environment[9]. This added layer of instability limits the current state of microbial biodesign and biases our understanding towards model organisms, even though more suitable hosts may exist for a given application[10–12]. Overall, the chassis-effect constrains the design-build-test cycle by demanding costly repetitions of trial-and-error experimentation. There remain major knowledge gaps as to which cellular processes underpin species-specific chassis-effects. Closing this knowledge gap undoubtedly enables more efficient engineering of novel hosts and consequently shifts the current engineering dogma towards the notion that the host itself can be considered a part to tune circuit functions[7,8], which is a paradigm shift towards a new understanding of different microbial species and their unique genomes as customized hardware for biodesign[13].

Previous efforts to address this knowledge gap has shown that device performances between bacterial hosts can be better explained by the differences in the growth physiology of hosts rather than genomic or phylogenetic relatedness[14]. In turn, the available physiological states for a given host are ultimately shaped by those functions encoded in their genomes, and perhaps more importantly, the expression of gene products that control cellular physiology. Deeper investigation into the gene expression responses towards

the activity of a heterologous genetic circuits are needed to uncover insight into which cellular processes underpin observable chassis-effects. Designing such a study is difficult, as it requires a suit of microbial hosts capable of operating an identical genetic device, produce an observable chassis-effect under the same growth conditions and share significant enough amount of genomic identity for their interspecies gene expression to be comparable. We developed a synthetic biology kit for the *Stutzerimonas* genus to specifically address this challenge. Recent re-examination of the *Pseudomonas* genus invigorated with high quality genome sequencing data led to the delineation of the clade into several novel genus[15,16], one such genus being *Stutzerimonas*[17,18]. Several members of the *Stutzerimonas* with sequenced genomes are available in culture collections and previous studies has highlighted the natural high transformation competence of *Stutzerimonas* spp[19,20]. Furthermore, many *Stutzerimonas* members have innate phenotypes that have garnered attention as potential microbial cell factories and/or bioremediation agents[21–24], making them appealing targets for domestication and biodesign applications.

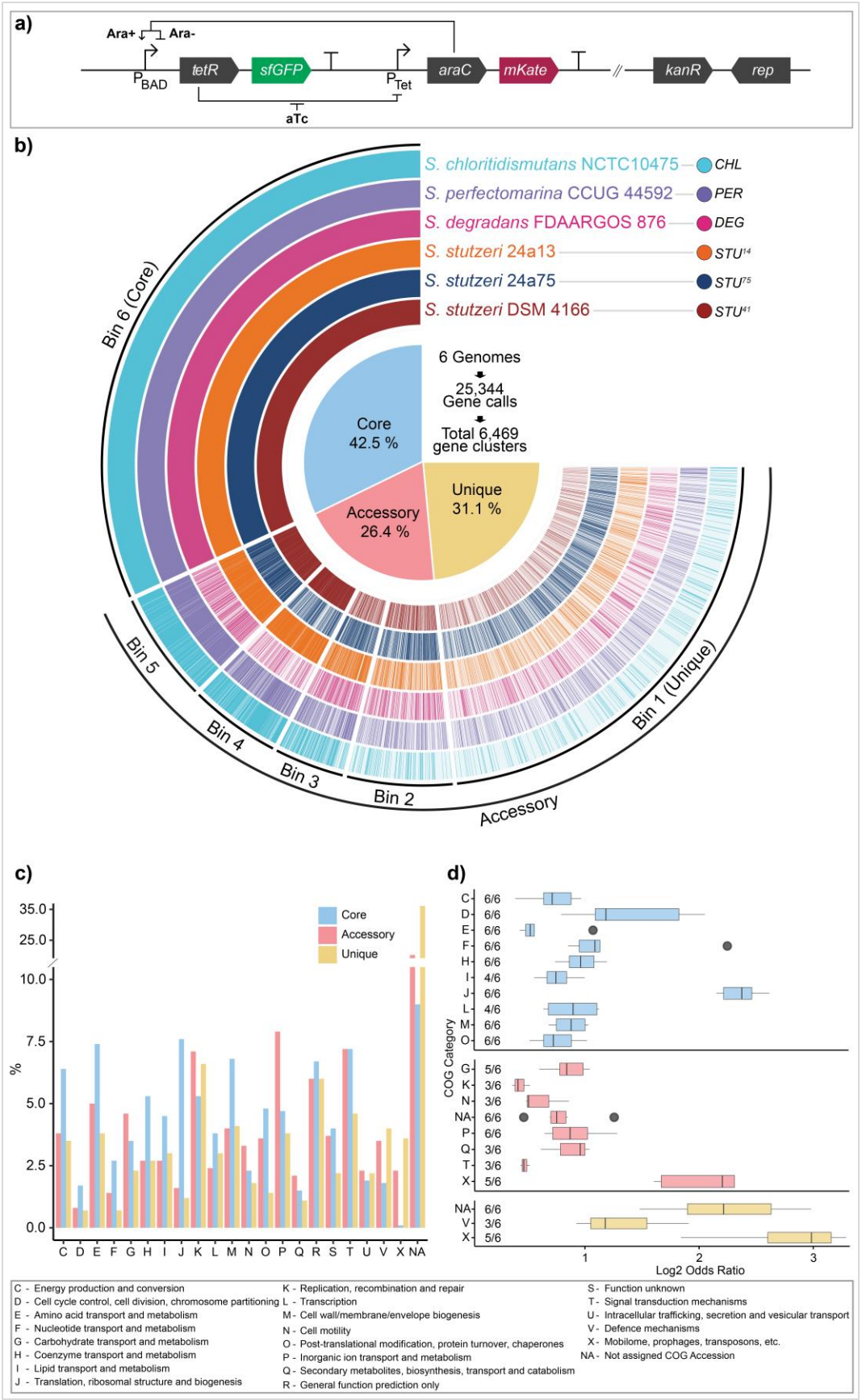
In this work, we transformed an inducible genetic inverter device into six *Stutzerimonas* hosts that share sufficient amount of genomic identify to form a quality pangenome[17], consisting of a core (i.e., orthologous genes shared between all hosts) and an accessory genome (i.e., orthologous genes shared only by a subset of hosts or genes unique to host). We comparatively quantified an observable chassis-effect in device performance within these closely related *Stutzerimonas* hosts and sequenced the global transcriptomes during different operation modes of the inverter. This investigative workflow enabled us to ask the guiding scientific question of whether the observed chassis-effect is more influenced by expression of conserved core genes that are fundamental to growth physiology and cellular housekeeping or by less conserved functions encoded within the accessory genome of our experimental platform. We were also able to ask if unique, species-specific gene expression patterns can serve as a concordant predictor of device performance. Our results show significant correlation between transcriptional response of shared core genes and inverter performance, as well as between growth physiology and inverter performance, suggesting the

mechanism in which core genetic elements contribute to the chassis-effect effect is through changes in basic cellular physiology related to growth.

RESULTS

A *Stutzerimonas* tool kit for pangenome-guided synthetic biology

Connecting structures and functions of multiple genomes to a measurable chassis-effect requires a standardized, broad-host-range genetic circuit. We therefore implemented a modified build from one of our previously described genetic inverters[14] (Fig. 1a). This inverter can be directionally induced (toggled) by anhydrotetracycline (aTc) and L-arabinose (Ara) and was cloned into plasmid pS5. The circuit features two inducible promoters, pBAD and pTet, with each promoter regulating the expression of the other's cognate transcription factor as well as a fluorescent reporter protein in a bicistronic manner. Fifteen *Stutzerimonas* hosts were screened for plasmid transformability and inverter operability, in which ten were successfully transformed via electroporation with a pBBR1-KanR backbone vector (BB23). Six of these ten were chosen for further study, which are as follows: *Stutzerimonas chloritidismutans* NCTC10475 (*S. chloritidismutans*)[25], *Stutzerimonas perfectomarina* CCUG 44592 (*S. perfectomarina*)[18], *Stutzerimonas degradans* FDAARGOS 876 (*S. degradans*)[18], *Stutzerimonas stutzeri* 24a13 (*S. stutzeri* 24a13)[26], *Stutzerimonas stutzeri* 24a75 (*S. stutzeri* 24a75)[26] and *Stutzerimonas stutzeri* DSM 4166 (*S. stutzeri* DSM 4166)[27,28].


a)
Ara+
Ara-
P BAD
tetR
sfGFP
P Tet
araC
mKate
kanR
rep
aTc
b)
S. chloritidismutans NCTC10475
CHL
S. perfectomarina CCUG 44592
PER
S. degradans FDAARGOS 876
DEG
S. stutzeri 24a13
STU14
S. stutzeri 24a75
STU75
S. stutzeri DSM 4166
STU41
6 Genomes
25,344
Gene calls
Total 6,469
gene clusters
Core
42.5 %
Accessory
26.4 %
Unique
31.1 %
Bin 6 (Core)
Bin 5
Bin 4
Bin 3
Bin 2
Bin 1 (Unique)
Accessory
c)
Core
Accessory
Unique
%
C D E F G H I J K L M N O P Q R S T U V X NA
d)
COG Category
Log2 Odds Ratio
C - Energy production and conversion
D - Cell cycle control, cell division, chromosome partitioning
E - Amino acid transport and metabolism
F - Nucleotide transport and metabolism
G - Carbohydrate transport and metabolism
H - Coenzyme transport and metabolism
I - Lipid transport and metabolism
J - Translation, ribosomal structure and biogenesis
K - Replication, recombination and repair
L - Transcription
M - Cell wall/membrane/envelope biogenesis
N - Cell motility
O - Post-translational modification, protein turnover, chaperones
P - Inorganic ion transport and metabolism
Q - Secondary metabolites, biosynthesis, transport and catabolism
R - General function prediction only
S - Function unknown
T - Signal transduction mechanisms
U - Intracellular trafficking, secretion and vesicular transport
V - Defence mechanisms
X - Mobilome, prophages, transposons, etc.
NA - Not assigned COG Accession

119

Figure 1. The genetic inverter and pangenome of selected *Stutzerimonas* hosts. a) Schematic representation of Ara-aTc genetic inverter composition. In presence of Ara (Ara+), Ara-bound AraC upregulates its cognate promoter (pBAD), leading to sfGFP and TetR expression and in turn creates a distinct measurable fluorescent state and leads to the downregulation of the pTet promoter. In absence of Ara (Ara-), AraC functions as a repressor. The presence of aTc leads to mKate and AraC production. The two promoters thereby act antagonistically, where upregulation of one lead to downregulation of the other. b) Gene clusters identified from pangenome analysis of the six *Stutzerimonas* hosts arranged by number of hits in host genomes where Bin 6 means all six hosts contribute with at least one gene call to the gene cluster and Bin 5 means any combination of 5 hosts contribute with at least one gene call and so on. Bins 5 to Bin 1 are further grouped together as the accessory genome, with gene clusters belonging in Bin 1 distinguished as unique. Percentages indicates portion of the 6469 gene clusters assigned to the three frequency groups (core, accessory and unique). c) Percentage COG category composition of each frequency group. d) Enrichment analysis of COG categories within each frequency group by Fisher exact test. Number "*n*/6" indicates number of hosts in which COG category was found significantly (P-value < 0.05, Bonferroni correction) enriched within the frequency group. Only COG categories enriched in 3 or more hosts are shown. COG category description provided at bottom of figure. CHL = *Stutzerimonas chloritidismutans* NCTC10475; DEC = *Stutzerimonas perfectomarina* CCUG 44592; DEG = *Stutzerimonas degradans* FDAARGOS 876; STU[13] = *Stutzerimonas stutzeri* 24a13, STU[75] = *Stutzerimonas stutzeri* 24a75; STU[41] *Stutzerimonas stutzeri* DSM 4166.

Comparative pangenomics of the selected *Stutzerimonas* hosts revealed almost equal sizes between the core and accessory genomes, propitiously setting the stage for investigating whether the measurable chassis-effect can be attributed to differential expression from the core or accessory genomes and which specific functions covary with host-specific device performance. Pangenome analysis of *Stutzerimonas* host genomes was performed using Anvi'o[29], leading to a total of 25,344 gene calls functionally annotated using the 2020 clusters of orthologous genes (COGs) database (Fig. 1b). Under the most sensitive settings, these gene calls were grouped into 6,469 gene clusters or "pangenomic orthologous groups". A gene cluster is grouped into "core" or "accessory" frequency group, depending on the number of genomes the gene cluster occurs in. Core gene clusters were defined as the 42.5 % (2,751) of gene clusters that had hits across all

hosts. In contrast, the "accessory" gene clusters made up the remaining 57.5 % of gene clusters that had hits in five or fewer genomes. Within the accessory genome, we further distinguish gene clusters exclusive to a single genome, referred to as "unique" gene clusters, which makes up 31.1 % (2,013) of total gene clusters. The number of host-specific gene calls ranged from 3,732 in *S. degradans* to 4,457 in *S. stutzeri* 24a13, with all hosts sharing on average 67.2 ± 4.4 % of genes, with 25.1 ± 6.3 % and 8.2 ± 3.6 % of gene clusters categorized as accessory and unique, respectively. 23.1 % of gene clusters were not assigned to any COG category (assigned to "NA" group) (Fig. 1c), with an additional 3.6 % and 4.9 % assigned to COG category S (Function Unknown) and R (General Prediction Only). 81 % of unassigned gene clusters belong in the accessory genome, with 54.7 % of these being unique gene clusters. This disproportionate number of unassigned gene clusters between the core, accessory and unique groups is consistent with previous bacterial pangenome studies[17] and is theorized to be due to the accessory genome tending to house genes that confer a specific advantage to the organism within its niche environment and therefore might only be expressed under certain conditions, making it difficult to study such under cultivation conditions that have been standardized across multiple species.

The seemingly uneven distribution of COG categories across the three frequency groups prompted an investigation of whether certain COG categories are overrepresented (Fig. 1d). The Fisher's exact test (*P*-value > 0.05, with Bonferroni Correction) revealed that within the core genome, COG categories consistently enriched for (i.e., significantly enriched in ≥ 3 species) are categories associated with housekeeping functions such as category C (Energy Production and Conversion), D (Cell Cycle Control, Cell Division and Chromosome Partitioning), E (Amino Acid Transport and Metabolism) and O (Post Translational Modification, Protein Turnover and Chaperones). Meanwhile, COG groups G (Carbohydrate Transport and Metabolism), P (Inorganic Ion Transport and Metabolism), Q (Secondary Metabolites Biosynthesis, Transport and Catabolism) and K (Transcription) are overrepresented in the accessory genome, suggesting the presence of metabolic pathways that confer unique nutrient assimilation

capabilities. Few COG categories were found to be consistently overrepresented within the unique genome, mainly due to majority of genes being of unknown function, but the few are category X (Mobilome: Prophages and Transposons), which hints towards past viral infection events, and V (Defense Mechanisms). Other prokaryotic pangenome studies report similar pattern of enriched COG categories in their defined shared and accessory genomes[30,31]. The distinct functional identity of genomic sub-groups gives merit to the hypothesis that expression of genes among the core or accessory groups may uniquely contribute to the chassis-effect.

The chassis-effect is observable between closely related *Stutzerimonas* hosts

Comparative measurements representing the performance of our engineered genetic inverter operated by each host under a standardized environment revealed a clear chassis-effect. Performance within *Escherichia coli* DH5α (*E. coli*) was also quantitatively compared as a positive control reference from a model organism. Induction response dynamics were characterized by fitting the Hill function $[(\beta x^n / (K^n + x^n)) + C]$ to normalized induction curves (Fig. 2a-d). Parameters C is the baseline output at 0 inducer concentration, β is the max output level at saturating input levels, K represents the sensitivity of the system to inducer input as well as the input responsive range, and the Hill coefficient n reflects the steepness of the response (varying from step-like or dosage dependent). These parameters collectively quantify interactions between inducers (aTc and Ara), their respective transcriptional factors (TetR and AraC) and the responsive operons; thereby quantitatively describing device performance. Markedly different performance profile is observed depending on the host context the genetic inverter operates from – i.e., a strong chassis-effect. For instance, the observed K_{aTc} values range from 1.45 to 34.24 nM aTc among the *Stutzerimonas* hosts (Fig. 2a, b), suggesting host-specific factors affecting the intracellular aTc concentration and/or different levels of TetR repressor. *E. coli* exhibited the highest sensitivity to aTc, with a K_{aTc} value of 0.48. The K_{Ara} metrics were relatively more uniform across hosts, with only an overall 3.8-fold difference between the smallest and largest K_{Ara} values (Fig. 2c, d). We define an additional metric, $DR_{T\text{-}Ara}$ and $DR_{T\text{-}aTc}$ (theoretical dynamic

range) as the ratio of the estimated β and empirical C expressed as a fold-change value, which informs the theoretical possible fold-change difference in output as measured from respective reporter proteins. At saturated inducer concentrations, the highest β_{Ara} achieved among the *Stutzerimonas* hosts was *S. perfectomarina* at 12,500 RFU, which also achieves the highest $\text{DR}_{\text{T-Ara}}$ value of 24.0. Overall, *E. coli* had the lowest β_{Ara} at 750 RFU. The range of observed β_{aTc} and $\text{DR}_{\text{T-aTc}}$ values are much lower, from 2,300 RFU to 5,900 RFU, and 2.0 and 3.8, respectively.

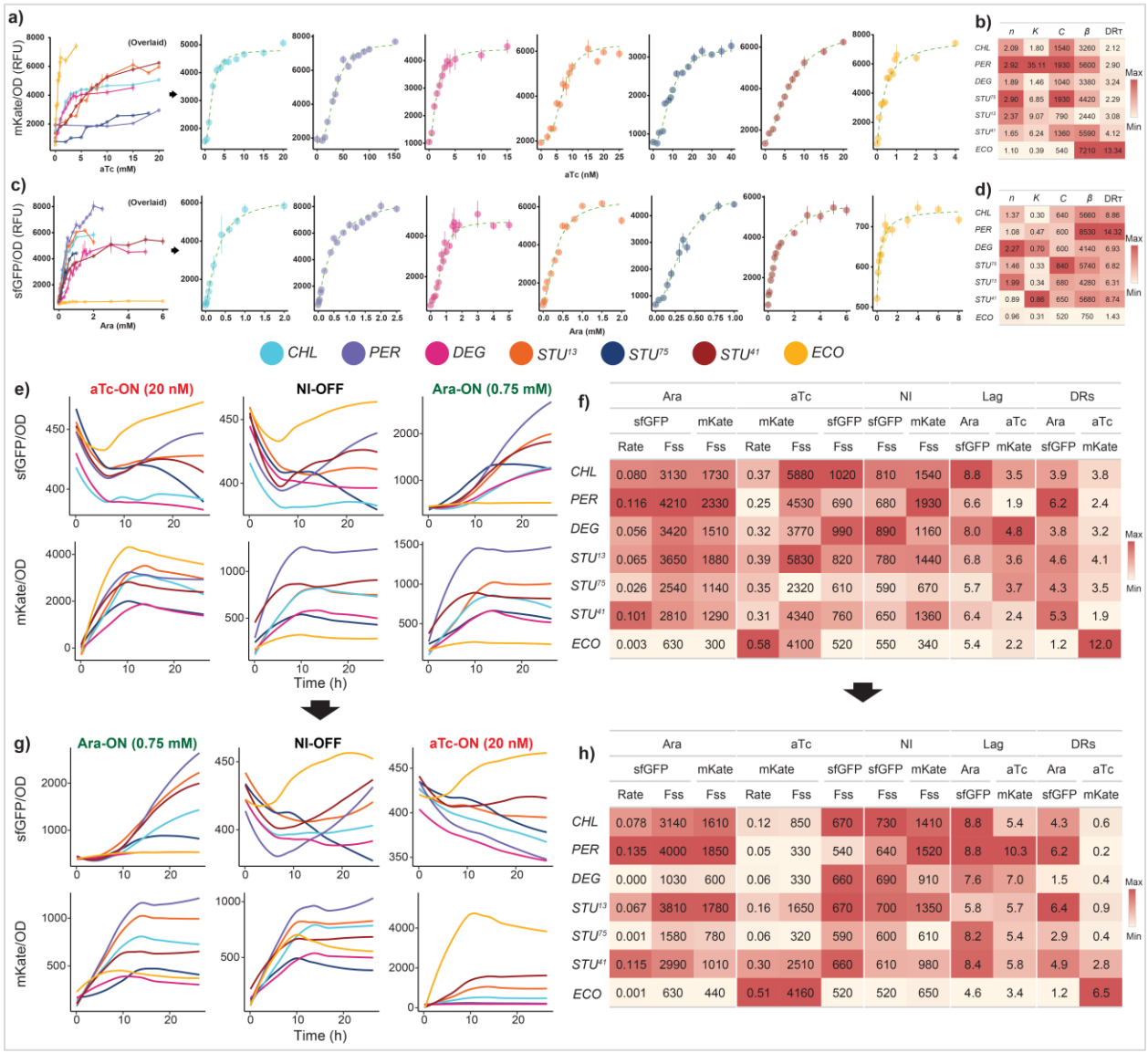


Figure 2. The chassis-effect is observed through the measurable performance of the genetic inverter between closely related *Stutzerimonas* hosts. a) aTc induction curves with left-most plot showing overlaid induction curves

up to given inducer concentration, followed by individual induction curves with scaled axis. Hosts are color coded, error bars indicate standard error of the mean, n = 8. b) Estimated Hill parameters from aTc induction curves. Color scale relative to each column. c) Ara induction curves and d) estimated Hill parameters. e) OD_{600} normalized fluorescence dynamics of one of three toggle assays with induction scheme 0.75 mM Ara and 20 nM aTc. Initial OFF cells were diluted to respective induction states. f) Estimated fluorescence metrics from fluorescence in e) across induction state and fluorescence output type. g) Fluorescence dynamics of toggled cells diluted to respective opposite inducer and h) estimated fluorescence metrics. NI = no induction, Fss = late phase steady-state fluorescence, Rate = max specific rate. DRs = Specific dynamic range.

We next performed a toggle assay to demonstrate the invertibility of the device and to further quantify the chassis-effect by the different degrees of hysteresis experienced by each host (Fig. 2e-h). Initial OFF cells (grown in absence of inducer) were diluted into media with inducer to prompt Ara-ON or aTc-ON states, respectively (Fig. 2e) and performance metrics from normalized sfGFP and mKate fluorescence curves was determined across induction states. These metrics (Fig. 2f) include the maximum steady/state fluorescence at late growth phase (F_{ss}), maximum rate of fluorescence (Rate) at exponential phase, specific dynamic range (DR_S) and lag time (Lag). Each of these metrics have different biological implications, depending on the induction state. For example, the fluorescence output under OFF conditions is a measure of the inverter's unbiased output level – i.e., the host-specific background fluorescence. Meanwhile, fluorescence outputs from cells grown under presence of an antagonistic inducer (e.g., sfGFP output in presence of aTc) indicate the deficient transcriptional control (leakage). The range of observed fluorescence curves and estimated performance metrics further solidifies the existence of the chassis-effect. Inversion between states was controlled by washing cells twice and diluting them into to media containing respective opposite inducer (Fig. 2g, h).

Results from the toggle assay showed that upon being toggled from Ara-ON to aTc-ON, multiple hosts experienced an attenuated mKate output, to different degrees. For instance, the DR_{S_aTc} values of all hosts other than *S. stutzeri* DSM 4166 all drop below 1 when diluted from Ara-ON to aTc-ON, meaning their

induced mKate output becomes lower than their baseline output. This decreased output was not observed when diluted from OFF to aTc-ON and can therefore be attributed as a measurable hysteresis effect. Two additional toggle assays with induction schemes 0.25 mM – 40 nM aTc and 0.375 mM Ara – 5 nM aTc was performed (Figure S1), and for *S. chloritidismutans*, *S. degradans* and *S. stutzeri* DSM4166, the mKate output attenuation was relieved under induction schemes with decreased Ara concentration but hosts the other three hosts experienced consistent attenuation across toggling schemes. Since cells were washed twice and diluted, the mechanism in which mKate output is attenuated is likely not due to residual extracellular Ara concentrations. This result suggests the invertibility (input/output logic) of the inverter is dependent on past induction states and that dynamics of this hysteresis-effect varies between hosts based on intracellular molecular physiology.

Species-specific physiology responses to the genetic inverter

A consistent growth inhibition was observed when comparing wild type hosts against their engineered genotypic counterparts and across induction states (Fig. 3). Growth of hosts operating the genetic inverter was measured simultaneously during the toggle assay from which growth physiology was characterized (Fig. 3a-d). The reference host (*E. coli*) demonstrated an overall more rapid growth on LB media compared to all *Stutzerimonas* hosts. The addition of BB23 backbone lead to decreased specific growth rates of all strains compared to their respective wild type counterparts, with *S. stutzeri* 24a75 exhibiting the most reduced growth rate along with *S. stutzeri* 24a13 and *E. coli*. (Fig. 3e, f). This is expected due to the added burden of maintaining the vector backbone in presence of kanamycin. *E. coli* and *S. chloritidismutans* exhibits the greatest additional growth burden upon addition of the inverter device into the backbone (complete pS5 plasmid) under NI conditions. The added burden is likely due to larger plasmid payload[32] and possible leakage expression of the device. However, some hosts experience close to zero additional growth burden, and while *S. chloritidismutans* had relatively high amount of leakage expression (sfGFP and mKate output in absence of inducer) (Fig. 2c, right), *E. coli* exhibited the lowest leakage expression,

suggesting the mechanism in which the backbone vector and device imposes growth inhibition on its host varies between hosts. Induction of the genetic inverter exacerbated growth inhibition (Fig. 3e, f), which also varied within hosts. For instance, the growth rate of *S. stutzeri* 24a75 was almost halved when induced with Ara when compared with NI condition but showed little response when treated with aTc. Overall, these results exemplifies that the genetic inverter´s mode of operation uniquely affects the growth of each host.

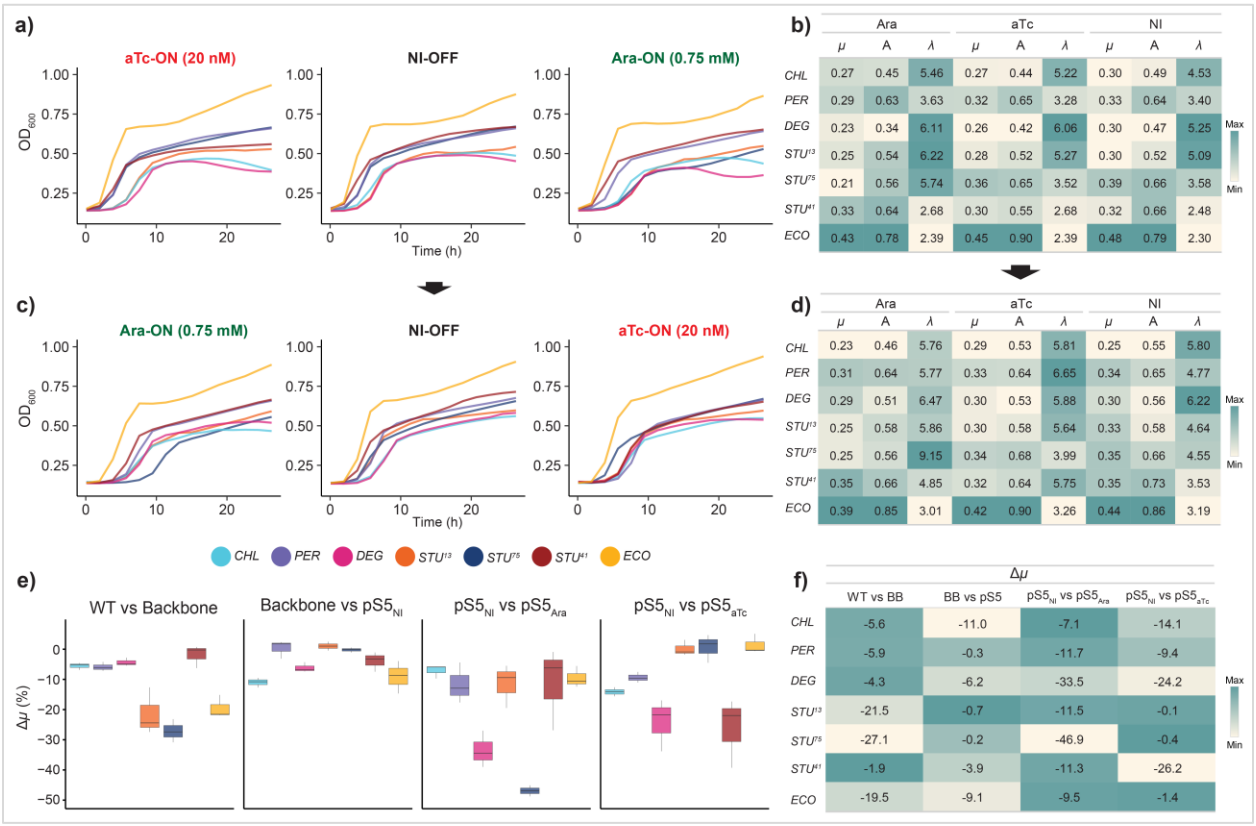


Figure 3. The growth dynamics are uniquely affected as a result of host-specific operation of genetic inverter. a) Growth curves of initial OFF cells diluted to respective induction states. Hosts are color coded. b) Estimated growth metrics for each host from growth curves in a). Color scale is relative to each column. c) Growth curves of toggled cells diluted to respective opposite inducer to toggle induction state and d) corresponding estimated growth metrics. e) Growth difference between host genotypes and/or induction state captures by the $\Delta\mu$ metric, defined as relative percentage change in growth rate. WT = wild type, BB23 = pBBR1 and KanR cloning vector. $pS5_{NI}$ = pS5 plasmid in

absence of inducer. $pS5_{Ara}$ = pS5 plasmid in presence of Ara (0.75 mM), $pS5_{aTc}$ = pS5 plasmid in presence of aTc (20 nM). NI = no induction; μ = max specific growth rate; A = carrying capacity; λ = lag time.

The unique intracellular molecular environment within each host is ultimately shaped by their respective genomes, and perhaps more importantly, the expression pattern of their gene products. We therefore extracted total RNA from cells induced in both directions for mRNA sequencing to determine their transcriptome profiles. By augmenting our comparative transcriptome analysis with pangenomic insight, we can discern the impact that the core and accessory genome has on the observed host-specific inverter performance.

Unique transcriptional patterns arise from both presence and operation of an engineered genetic circuit

Global transcriptome analysis revealed marked variability in gene expression profiles between hosts operating the engineered genetic inverter. The difference in response was observed in terms of magnitude (number of differentially expressed genes or DEGs), the functional composition of DEGs and direction of regulation of certain gene clusters shared among hosts´ core genomes. These unique transcriptomic profiles support our hypothesis that differences in gene expression from core and/or accessory genomes uniquely correspond with differences in genetic inverter performances.

Cross-species comparison of DEG profiles was performed by pooling all read counts mapped to gene calls within a gene cluster, on the basis that genes within the same gene cluster are inferred to be highly similar. Differential expression of mRNA encoded from the genetic inverter confirmed its programed operability (Fig. 4). Cells of each host that were treated with Ara showed a higher proportion of reads (measured by transcripts per million, TPM) mapped to *tetR* and *sfGFP* compared to *araC* and *mKate*, and vice versa for aTc induced cells (Figure S2). Differential gene expression analysis of Ara against aTc treated hosts reveals

significant upregulation of *tetR* and *sfGFP* genes and downregulation of *araC* and *mKate*, in accordance with the design of the inverter. The degree of response differed greatly between hosts. Log2 fold-change values of *tetR* ranged from 4.9 (*S. chloritidismutans*) to 8.2 (*S. degradans*) and ranged from 1.1 (*S. perfectomarina*) to 4.1 (*S. stutzeri* 24a75) for *sfGFP*. These results support the observed chassis-effect via mRNA abundances from genes encoded within the engineered genetic inverter. The consistent downregulation of the *kanR* gene in Ara induced cells could be due to higher expression of *kanR* in aTc induced cells because of transcriptional readthrough when transcribing from the pTet promoter. A relatively large difference in TPM values was also observed between polycistronic repressor-reporter pairs (Figure S2).

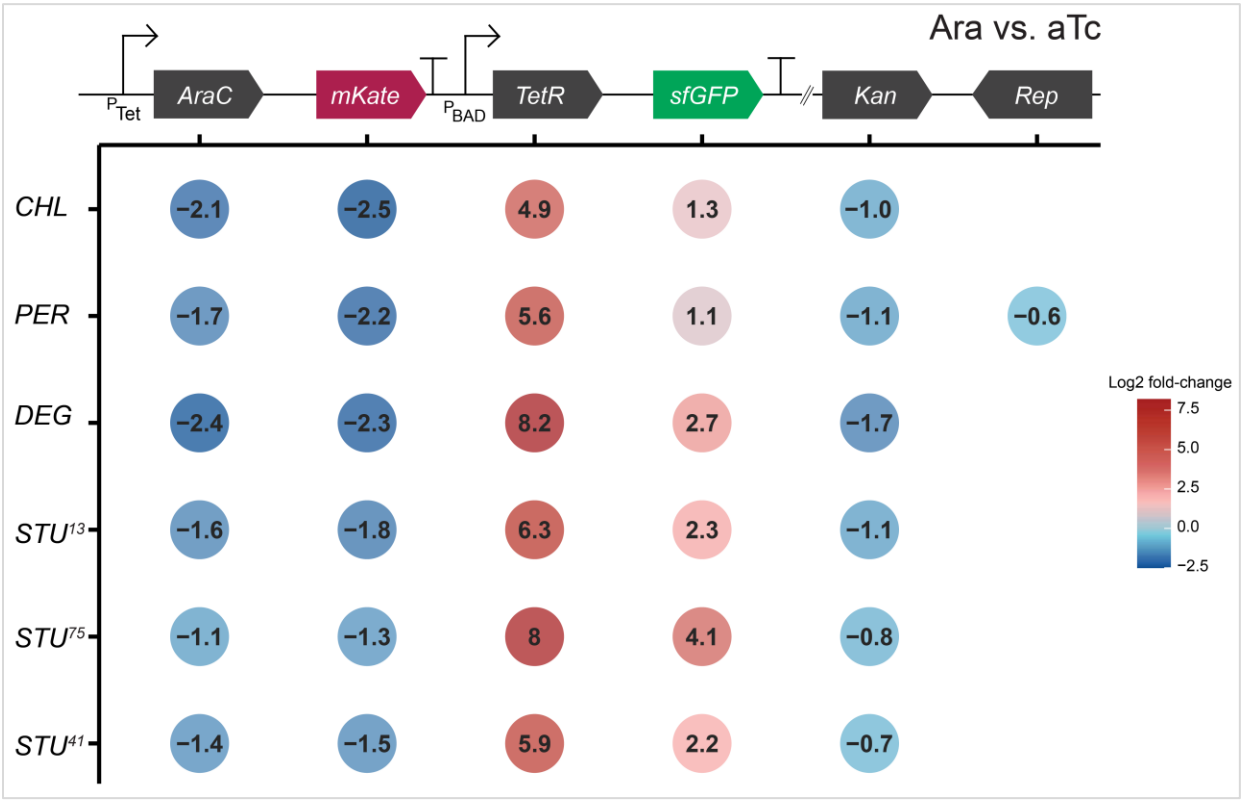


Figure 4. The chassis-effect is measurable at the differential response of genes designed into the engineered genetic inverter. Log2 fold-change values of the six genes encoded in the pS5 plasmid between hosts comparing Ara against aTc induced cells. Empty position indicates non-significant differential expression.

The mRNA abundance of genes encoded by the genetic inverter intuitively play a major role in influencing the observed chassis-effect, but these heterologous gene products are not compartmentalized away from native cellular elements. We were therefore prompted to also explore the hosts' global transcriptome response to better understand how expression of genes from the core and accessory genomes might be concordant with the chassis-effect. Our results reveal a clear diversity in the global differential gene expression responses between hosts when comparing Ara against aTc induced cells, indicating that genome encoded functions respond differently depending on how the engineered genetic inverter is being operated (Fig. 5a, b). Differential expression analysis with a *P*-value threshold of 0.05 and log2 fold-change threshold of 0 leads to a total of 4,825 significantly differentially expressed genes (DEGs) distributed across hosts, with 3,995 DEGs belonging to the core group distributed among 1672 gene clusters. We observed that each host significantly expressed only a subset of shared core gene clusters, with 63 % of core gene clusters being expressed by at least three hosts. Among significant DEGs, the direction of regulation and the strength of response (captured by log2 fold-change value) within a gene cluster differs (Fig. 5c). The number of DEGs is unevenly distributed between the hosts, varying from 589 (*S. chloritidismutans*) to 1106 (*S. stutzeri* 24a75) (Fig. 5d) meaning the operation of the genetic inverter (*i.e.*, the user defined induction state) incites a greater response in some hosts than others. We note that the differences in DEG response between states cannot be entirely attributed to the expression of the inverter alone, but also from the presence of inducer compounds Ara and aTc. However, the addition of induction compounds is inherently coupled to the function of the inverter by design and can therefore include responses to inducers as a contributor to the chassis-effect.

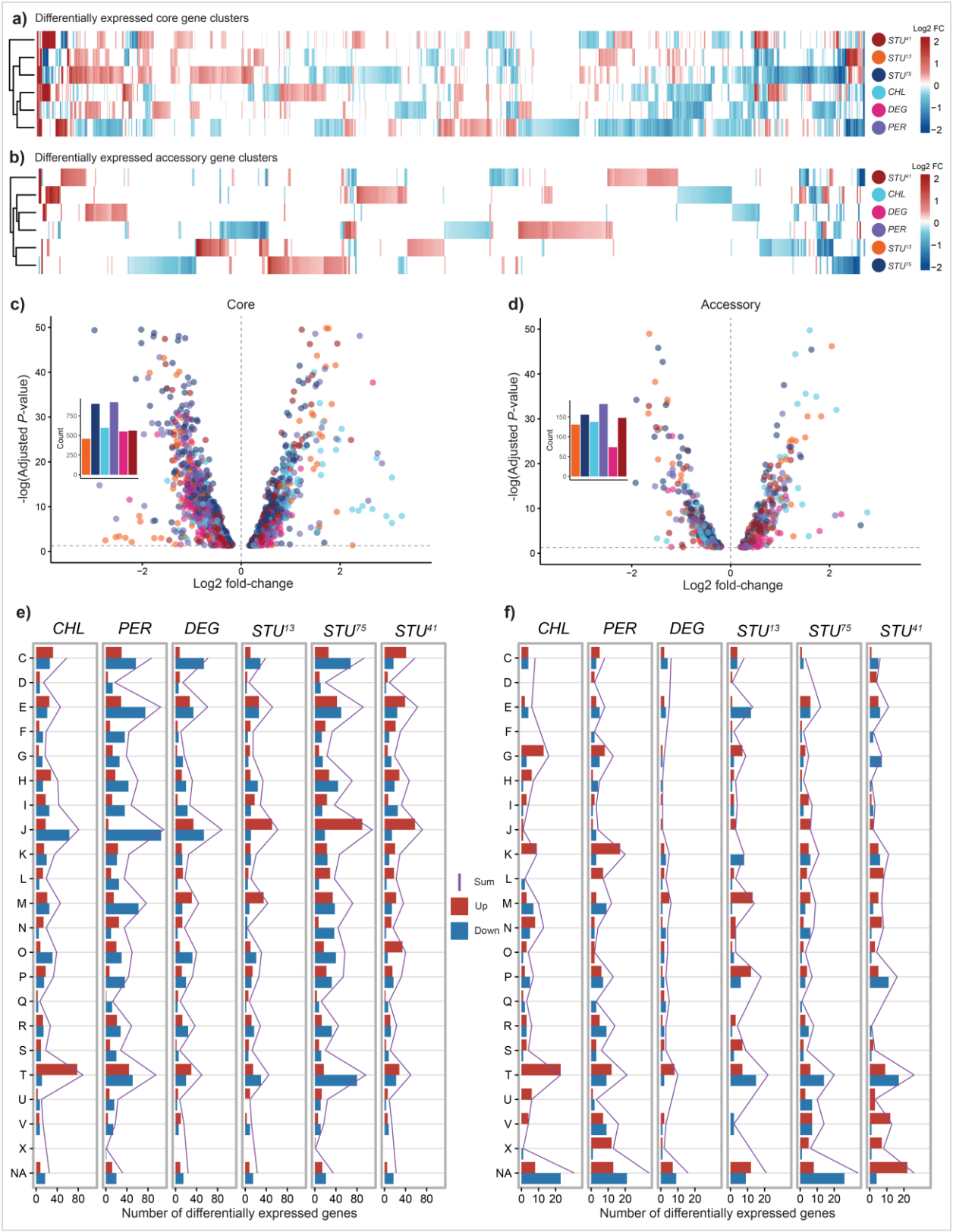

a) Differentially expressed core gene clusters
Log2 FC
STU41
STU13
STU75
CHL
DEG
PER
b) Differentially expressed accessory gene clusters
c) Core
d) Accessory
-log(Adjusted P-value)
Log2 fold-change
Count
e)
f)
Sum
Up
Down
Number of differentially expressed genes

Figure 5. Global differential gene expression analysis of Ara against aTc induced cells reveals diverse transcriptional profiles as a result of inverter operation. Log2 fold-change values of differential expressed a) core and b) accessory gene clusters significantly expressed by at least one host (*P*-value < 0.05). White bar indicates non-significant expression in host. For accessory genome, white bar indicates either non-significance or gene cluster has no hits for that host. Volcano plots visualizing log2 fold-change distribution of significantly DEGs in c) core genome and d) accessory genome. Inset bar charts show the number of differentially expressed genes for each host. The distribution of DEGs across COG categories across hosts are shown for both e) core and f) accessory. DEGs are further grouped into upregulated (red bars) and downregulated (blue bar) gene clusters within each COG category. The purple line denotes the sum of number of DEGs.

We next examined the functional profile of the core and accessory DEGs by assessing their distribution among the COG categories. The core DEGs of hosts shows a surprisingly uniform distribution across the 22 COG categories while a more varied response pattern in the accessory genome is observed. 45.3 ± 1.5 % of the DEGs can be found in categories C, E, J, M, T alone (Fig. 5e). Meanwhile in the accessory genome, COG categories P, K, T and NA (unassigned) make up 43.0 ± 2.6 % of DEGs (Fig. 5f), with NA gene clusters making up 17.1 ± 2.3 % alone. We observed difference in the composition of up and down regulated genes between hosts within functional groups (especially J, T, and M in the core group) suggesting varied responses to inverter activity. The less uniform transcriptomic responses from accessory genomes are attributed to gene clusters shared by only a subset of species. In the accessory genome, larger numbers of DEGs were concentrated in category T (Signal Transduction) for all hosts, indicating unique response patterns occurring within each host as a result of inverter activity. A larger variation in distribution of DEGs was observed within a COG category as well. For instance, 31.3 % of DEGs in category K occured in *S. perfectomarina* alone, further exemplifying the diversity in gene expression response pattern. We. note that the majority of DEGs are unassigned (NA), meaning with improved gene annotation, the transcriptional profile of the accessory DEGs could change substantially. The accessory genome, comprising genes with specialized functions shared by only a subset, can be a strong source for genes contributing to an observed

chassis-effect. However, differential expression of functions encoded within the core genome often belong to the major carbon, nitrogen and energy metabolism, cell division and housekeeping genes, which can more intuitively underpin the chassis-effect, given that majority of DEGs stem from the core genome.

Differential expression of core genes is concordant with the chassis-effect

Procrustes Superimposition (PS) analysis revealed significant concordance between the host specific expression of the core genome and genetic inverter performances. In contrast, there were no significant correlations between the accessory genome and device performances. Hence, the genome encoded functions most responsible for the observed chassis-effect arise from genes shared commonly between hosts. PS analysis compares the similarity between two configurations on a coordinate, and when the distances between points the configuration carry information of (dis)similarity, such as points on principal components, PS analysis can be used to determine whether similarities between two datasets correlate[33,34]. PS analysis was therefore performed to formally test for concordance between the PCA ordinated configurations of the hosts in terms of their similarity of genetic inverter performance and expression of their genome. This data exploration was done in a hierarchal manner, by first considering all significant DEGs, and then dissecting the transcriptome into core and accessory. Significancant correlations were measured when comparing the configurations from inverter performance and when taking into consideration all DEGs (Fig. 6a, P-value = 0.044, m^2 = 0.344). Upon splitting the DEGs into respective core and accessory groups, we found that it is in fact the core genome that is responsible for the significance observed previously (P-value = 0.034, m^2 = 0.324) (Fig. 6b, c). This result supports the following notions, that the observed chassis-effect can be explained by the differential gene expression response between hosts and that hosts with more similar expression pattern of their shared core genome also have significantly more similar performances. Hence, our given set of species and their shared gene clusters, core genes commonly associated with housekeeping functions, central carbon, and energy metabolism, are major biological drivers of the chassis-effect. Additionally, PS analysis applied to the captured growth metrics

show significant correlation with inverter device performance (Figure S3), corroborating our previous result that observed differences in the physiological state of hosts can be used to potentially predict device performance[14].

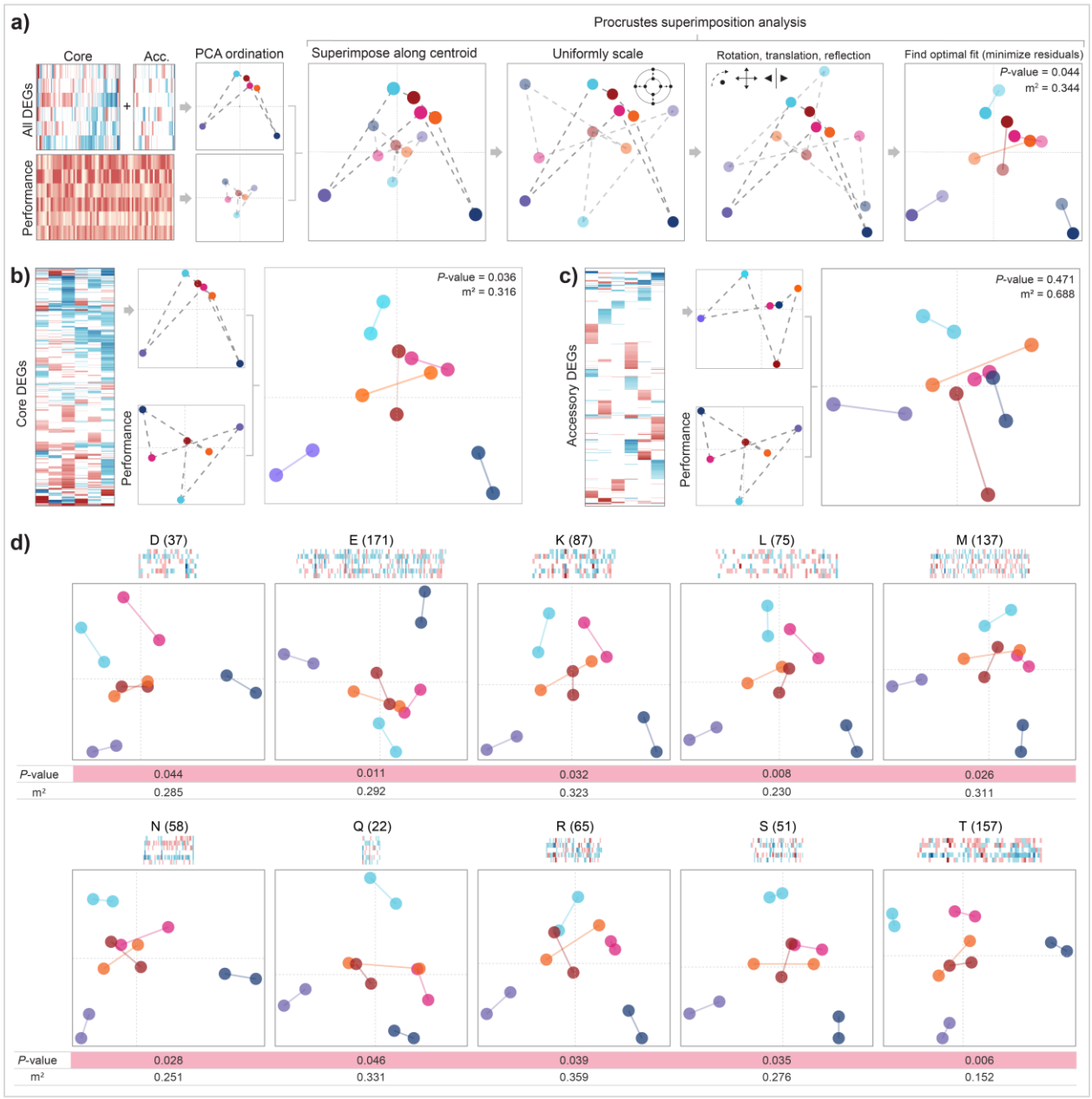


Figure 6. Procrustes analysis reveals significant concordance between similarity in inverter performance and similarity in core genome response between hosts. a) Procrustes superimposition analysis comparing hosts in terms of their total significant differential gene expression response against inverter performance metrics. The key steps in

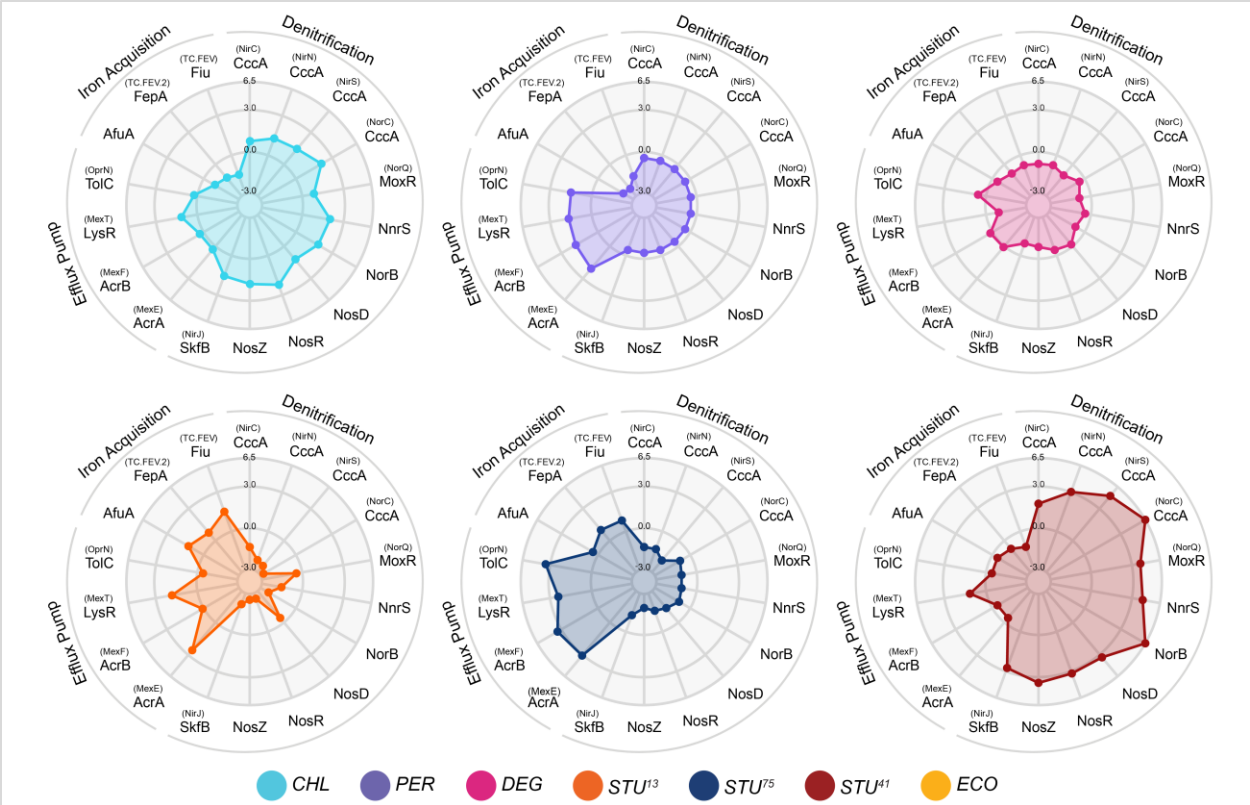
PS analysis are schematically illustrated. Performance metric dataset and differential gene expression dataset are first projected onto ordinate space via PCA, then the configurations are compared through PS analysis, which involves centering, scaling, and transforming the two projections with the goal of minimizing the sum of squared vector residuals (the m^2 statistic) between each respective point (host). Significance of the obtained statistic is determined through a permutation method. Dashed gray lines between points have been added to visualize an arbitrary "configuration" formed by each dataset, which connects each point in the following arbitrary order "-*CHL*-PER-*DEF*-*STU*^{13}-*STU*^{75}-*STU*^{41}-". b) Procrustes analysis comparing inverter performance against significantly differentially expressed b) core and c) accessory genome. d) Sub-groups of core genome grouped by COG category that were significantly correlated against inverter performance as determined by PS. Number in sub-plot headers indicates number of genes clusters within each group along with log2 fold-change profile of gene clusters. For gene clusters assigned to multiple COG categories, it was counted as belonging to each of those COG categories.

We next interrogated the specific cellular functions that are most concordant with the observed chassis-effect by dissecting the differentially expressed core and accessory genomes into their functional categories in a second iteration of Procrustes analysis. Ten of the twenty-three COG groups were found to be significant (Fig. 6d). Category E (Amino acid Transport and Metabolism) is especially intuitive because the main carbon and energy source provided from LB media are amino acids (not sugars) and different catabolic strategies for these will in turn cause different growth phenotypes. Categories R (General Function Prediction Only) and S (Unknown Function) were also significant, suggesting that genetic elements of unknown function are contributing strongly to the observed chassis-effect.

Genes clusters involved in denitrification and efflux pumps are highly responsive to genetic inverter activity within *Stutzerimonas* spp.

A data-driven investigation of the top 100 most differentiated gene clusters between host cells revealed that gene clusters involved in denitrification, iron acquisition and membrane-bound transport proteins are among the most highly differentially expressed gene clusters in response between Ara versus aTc treatment

applied to the *Stutzerimonas* hosts (Fig. 7). This was determined by ranking gene clusters by a metric that takes into consideration the number of hosts significantly expressing each gene cluster, the sum of absolute log2 fold-change values and the combinatorial sum of absolute differences in log2 fold-change values between species (Table S1). To add confidence in the functional annotation of the gene clusters, we supplemented the COG annotation with annotation using the KEGG Orthology (KO) database[35] because COG accessions are (in some cases) only general descriptions of protein functions or families. For instance, the assigned COG accession for the four highly ranked gene clusters IGC_00001064, GC_00001456, GC_00001627 and GC_00001727 describes all these gene clusters to encode for cytochrome c protein (CccA, COG 2010). The corresponding KO annotations for the four *CccA*-annotated gene clusters are cytochrome c55X (NirC, K19344), dihydro-heme d1 dehydrogenase (NirN, K24867), nitrite reductase (NirS, K15864) and cytochrome-containing nitric oxide reductase subunit c (NorC, K02305), respectively, which are all heme-containing enzymes belonging in the cytochrome c protein family[36,37]. Notably, every KO accession has a complementary COG accession entry in the KEGG Orthology database (https://www.genome.jp/kegg/ko.html), and the KO and COG annotations done independently by KofamKOALA and Anvi'o (for COG annotation) discussed here all match their respective database records, indicating that the COG and KO annotations are in agreement.



Gene cluster ID	Name	COG	COG accession description	KO	KO accession description	Path
GC_00001064	CccA (NirC)	COG2010	Cytochrome C, mono- and diheme variants	K19344	Cytochrome C55X, involved in nitrite reductase heme d1 cofactor biosynthesis	Denitrification
GC_00001456	CccA (NirN)	COG20101|COG3391	Cytochrome C, mono- and diheme variants | DNA-binding beta-propeller fold protein	K24867	Dihydro-heme d1 dehydrogenase	Denitrification
GC_00001627	CccA (NirS)	COG20101|COG3391	Cytochrome C, mono- and diheme variants | DNA-binding beta-propeller fold protein	K15864	Nitrite reductase (NO-forming) / hydroxylamine reductase	Denitrification
GC_00001727	CccA (NorC)	COG2010	Cytochrome C, mono- and diheme variants	K02305	Nitric oxide reductase subunit C	Denitrification
GC_00000037	MoxR (NorQ)	COG0714	MoxR-like ATPase	K04748	Nitric oxide reductase	Denitrification
GC_00001896	NnrS	COG3213	Nitric oxide response protein	K07234	Uncharacterized protein involved in response to nitric oxide	Denitrification
GC_00000522	NorB	COG3256	Nitric oxide reductase large subunit	K04561	Nitric oxide reductase subunit B	Denitrification
GC_00001169	NosD	COG3420	Nitrous oxide reductase accessory protein, contains tandem cash domains	K07218	Nitrous oxidase accessory protein	Denitrification
GC_00002351	NosR	COG03481|COG3901	Polyferredoxin naph transcriptional regulator nosr of nitric oxide reductase	K19339	NosR/NirI family transcriptional regulator, nitrous oxide reductase regulator	Denitrification
GC_00001565	NosZ	COG4263	Nitrous oxide reductase	K00376	Nitrous-oxide reductase	Denitrification
GC_00001908	SkfB (NirJ)	COG0535	Radical sam superfamily maturase, SkfB/NifB/Pqqe family	K24868	Heme d1 biosynthesis protein	Denitrification
GC_00000078	AcrA (MexE)	COG0845	Multidrug efflux pump subunit AcrA	K18298	Membrane fusion protein, multidrug efflux system	Efflux pump
GC_00000115	AcrB (MexF)	COG0841	Multidrug efflux pump subunit AcrB	K18299	Multidrug efflux pump	Efflux pump
GC_00002718	LysR (MexT)	COG0583	DNA-binding transcriptional regulator, LysR family	K18297	LysR family transcriptional regulator, MexEF-OprN transcriptional activator	Efflux pump
GC_00000605	TolC (OprN)	COG1538	Outer membrane protein TolC	K18300	Outer membrane protein, multidrug efflux system	Efflux pump
GC_00002662	AfuA	COG1840	ABC-type Fe3+ transport system, periplasmic component	K02012	Fe3+ transport system substrate-binding protein	Iron acquisition
GC_00001023	FepA (TC.FEV)	COG4771	Outer membrane receptor for ferrienterochelin and colicins	K16089	Outer membrane receptor for ferrienterochelin and colicins	Iron acquisition
GC_00002160	Fiu (TC.FEV)	COG4774	Outer membrane receptor for monomeric catechols	K02014	Iron complex outermembrane recepter protein	Iron acquisition

Figure 7. Gene clusters containing the most highly differentially expressed genes between hosts. Spider plots showing log2 fold-change data of most highly differentially expressed gene clusters between the hosts. Gene names in parentheses are names provided by KO annotation, all other gene names are provided by the annotated COG accession. In cases where the gene name provided by COG and KO match, only one gene name is shown.

Genes involved in denitrification were upregulated in *S. chloritidismutans* and *S. stutzeri* DSM 4166 and downregulated in *S. stutzeri* 24a13 and *S. stutzeri* 24a75 but showed little change in *S. decontaminans* and

S. degradans when comparing Ara against aTc induction states. This result was not anticipated and exemplifies that operation of engineered genetic devices can impact and/or be impacted by fundamental cellular functions seemingly unrelated to the heterologous transcription and translation networks. Denitrification is canonically described as an anaerobic process, but aerobic denitrification has been identified in numerous bacteria, many of whom have been identified as *Pseudomonas stutzeri* species[21,38,39]. The gene *napA* encodes for a nitrate reductase which catalyzes the important first step in the denitrification pathway. Only one gene cluster, GC_00001817, was annotated as *napA* by KO annotation, but this gene cluster was only significantly differentially expressed in *S. decontaminans* and ranked low. Genes involved in iron acquisition showed the opposite trend, being upregulated in *S. stutzeri* 24a13 and *S. stutzeri* 24a75 and downregulated to some degree in all other hosts. Denitrification activity requires bioavailable iron for heme biosynthesis[40], hence we expected the differential expression of the two pathways to instead positively correlate.

Other highly ranked gene clusters of note from the data-driven investigation are GC_00002718, GC_00000078 and GC_00000115. COG annotation inferred these gene clusters to encode for components of the AcrAB-TolC efflux pump. Meanwhile, KO infers them to be components of the MexEF-OprN efflux pump. Both pumps are broad-substrate Resistance-Nodulation-Division family transporters known to provide drug resistance[41,42]. Another high-ranking gene cluster was GC_00002718, annotated by COG as a transcriptional regulator part of the LysR family, but as MexT by KO, the latter being a LysR-type transcriptional regulator regulating the MexEF-OprN operon. The inclusion of the LysR regulator indicates that gene clusters GC_00002718, GC_00000078 and GC_00000115 indeed encode for a MexEF-OprN efflux pump, as AcrAB-TolC is regulated by the transcription factor AcrR, which instead belongs in the TetR family of transcriptional regulators[43]. Interestingly, hosts that exhibit downregulated or low change in expression of denitrification gene clusters were also the hosts with highest upregulation of the AcrAB-TolC/MexEF-OprN encoding gene clusters (hosts *S. decontaminans*, *S. stutzeri* 24a13 and *S. stutzeri*

24a75). Incidentally, Fetar, et al. 2011 has reported that nitrosative stress in the form nitric oxide accumulation is a direct inducer of MexEF-oprN expression[42], suggesting that hosts expressing the efflux pump could be a response to nitric oxide.

DISCUSSION

This study was driven by the overarching question as to whether closely related bacterial hosts can exhibit a strong chassis-effect when programmed with an identical engineered genetic inverter and how contributions from the core and accessory genomes might underpin this phenomenon. This was investigated using a supervised machine learning approach that employed dimension-reduction techniques combined with multivariate statistical analysis to determine which portions of the differentially expressed genome were concordant with the measured chassis-effect and then further trace these differences into specific genome encoded functions. In addition to our main finding, we report the successful engineering of several *Stutzerimonas* species. *S. degradans* is of interest with its known applications in bioremediation of contaminants[18] and *S. stutzeri* 24s75 is inferred to possess the *nifDKH* genes according to KEGG pathway mapping[44] and therefore has potential application as a biological nitrogen fixation agent to replace agrochemicals, similar to *S. stutzeri* DSM 4166, the latter in which has previously received much attention for its nitrogen fixation capability[27].

It is tempting to theorize that accessory and host-specific unique genes would be the leading cause of phenotypic distinction (and resulting chassis-effects), yet our findings empirically showed no such relations between inverter performance and the accessory genome. Instead, we observed concordance between transcriptional response of shared core genes and inverter device performance. In other words, hosts with more similar gene expression from their core genomes also have more similar performance, suggesting that the expression of the core genome, and in turn its protein products, have a greater impact on the cellular state. This holds especially true if the accessory genome consists of genes that are not expressed unless conditions call upon their expression, such as for unique biosynthetic gene clusters[45]. The core genome of our *Stutzerimonas* platform is enriched for genes involved in central carbon and energy metabolism as well as housekeeping-associated functions. Combined with the result that observed differential growth physiology between hosts was found significantly correlated with differential performance, the chassis-

effect may manifest through the redistributive flux of cellular resources and gene expression machinery unique to each host context because of genetic device maintenance and operation. This concurs with previous studies that has demonstrated how resource dynamics such as ribosome abundance[46] can affect genetic circuit performance and even parameterized resource competition to build predictive models[47,48].

Data-driven analyses showed that denitrification genes were responsive to heterologous expression of the engineered inverter. This indicates that engineered gene circuits can have unforeseeable and unspecific cross talk with fundamental cellular processes and that these events are context specific to individual hosts. If denitrification is in fact occurring, the source of nitrate in the media is unclear and would require a more targeted analysis on proteins and metabolic intermediates involved in denitrification pathways. A possible explanation is that our hosts can perform nitrification of ammonia to produce nitrate, with the source of ammonia being the by-product of amino acid catabolization from the oligopeptides in the LB media. But KEGG pathway mapping reveals that neither of two the genes involved in the canonical nitrification pathway (ammonia monooxygenase, *amo*, and hydroxylamine reductase, *hao*) are found in the genomes of our hosts (Supplementary Figure S4), suggesting a lack of nitrification capability. However, there has been numerous reports of other *Stutzerimonas* (reported as *Pseudomonas stutzeri*, now deprecated) capable of converting ammonia to nitrogen gas through heterotrophic nitrification and aerobic denitrification (HNAD) despite their genomes encoding neither *amo* nor *hao*[49–52]. These reported HNAD capable species are thought to perform HNAD through an undocumented pathway, or the enzymes catalyzing the reactions performed by *amo* and *hao* are significantly different enough in amino acid composition to escape current annotation algorithms. The same enigmatic HNAD process could be occurring for our *Stutzerimonas* species with increased denitrification activity. Denitrification a pragmatic phenotype due to its bioremediation potential[53], but bacteria capable of simultaneous HNAD are valuable as the cost of maintaining anaerobic conditions is reduced[21].

Given the complex interwoven network structure of cellular metabolism[54], it is impractical to postulate that the observable chassis-effect between a given set of hosts can be explained by a single or even a set of predictable genome-encoded functions without experimental insight. The examples from this study were dentification and efflux pumps, but given that it is host-context-specific, device operations in other species/strains may influence different cellular functions. Hence, deriving a mechanistic metabolic model detailing how these factors interact to bring about the produced output is likely intractable and may not hold true across different sets of hosts. Incidentally, with the advent of dimension reduction using multivariate statistical models in supervised machine learning, synthetic biologists can advance data-driven engineering strategies and bypass dependence on *a priori* mechanistic insight for biodesign applications[55,56]. In contemporary biological research, unveiling a mechanistic model that explains an observation is considered the pinnacle goal, but the goals of synthetic biologists more often prioritize the practical implementation of designed systems rather than knowledge-acquisition. Supervised machine learning aligns well with this goal[57].

We have developed new knowledge of how observable chassis-effects of an engineered microbial platform can be explained by genome encoded functions that are largely conserved and represent fundamental cellular processes. In addition, we uncovered a context-specific phenomenon that engineered genetic devices have unpredictable interference with metabolic functions that regulate cellular physiology. The evidence for these conclusions is corroborated by changes in species and even strain specific global transcriptome profiles and growth phenotypes, and the use of multivariate statistics to attribute how quantifiable chassis-effects are concordant with genome structure and function. This represents a genome-informed advancement within the field of broad-host-range biodesign which aims to lessen our reliance on model organisms so that we can better understand diverse microbial behaviors and use them in the blueprints of biodesign. The number of assimilated microbes available for use as industrial biotechnology platforms and biodesign engineering is increasing steadily[58–60], heralding the advancement towards broad-

host-range synthetic biology where synthetic biologists must explore not only the design space of genetic parts, but also the design space of host-chassis in order to optimize their engineered systems. As the field of biodesign progresses towards this new era, the constraints imposed by the chassis-effect will only become more relevant. It is therefore of high interest to develop novel hosts and data-driven predictive frameworks capable of using genomic insight to understand how chassis-effects might limit or in some cases expedite design-build-test cycles for the future´s biotechnology applications such as bioremediation of soil and marine contaminates and production of renewable agrochemicals.

MATERIALS AND METHODS

Species, cultivation, cloning, and transformation

Overview of species used in this study can be found in Table S2. The six *Stutzerimonas* hosts selected for further study were identified by verifying *rpoD* sequence through sanger sequencing[61] (Table S3). Cells were cultured in Lysogeny-Broth (LB) at 35 °C unless specified otherwise. BB23 backbone and pS5-carrying strains were cultivated in presence of 100 µg/mL kanamycin while wild types were grown without. Single colonies from streaked LB agar plates were picked to inoculate liquid media and incubated overnight with shaking to prepare overnight cultures. 199 µL of media was inoculated with 1 µL of overnight culture in black clear-bottom 96-well plates (Thermo Fischer, 165305) and sealed with Breath-Easy film (Sigma-Aldrich, Z380059). OD_{600}, sfGFP (Ex 485/ Em 515, gain 75) and mKate (Ex 585/ Em 615, gain 125) fluorescence was measured continuously using a Synergy H1 plate reader (Agilent Biotek, Serial Number 21031715) with continuous linear shaking (1096 cpm, 1 mm) at 9 mm read height. Working stock solutions of 1 M L-Arabinose (VWR, A11921) stock and 1 mM aTc (VWR, CAYM10009542) was prepared by dissolving powder in water and 70% ethanol respectively. Cloning was performed using *E. coli* DH5α, made chemically competent and transformed following the Inoue method[62]. *Stutzerimonas* species were transformed via electroporation method as previously described in Chan et al. 2023[14]. Primers used in this study can be found in Table S4. BB23 backbone was integrated into the BASIC Assembly format using from pSEVA231 as template with primers B_SEVA_F and B_SEVA_R. Plasmid pS5 was assembled in the Biopart Assembly Standard for Idempotent Cloning (BASIC)[63,64] environment as previously described[14]. Sequence and accession of pS5 components can be found in Table S5.

Induction assays

Overnight culture grown in absence of inducer was used to inoculated to media with various concentrations of aTc and Ara in 96-well plates. The normalized steady-state fluorescence at late growth phase (Fss)

averaged oved a time window of 6 – 12 hours was used as response variable of induction curves. In R (v4.3.1), Hill coefficient (*n*), activation coefficient (*K*) and max steady-state fluorescence output (β) was estimated by fitting the Hill function (1) using non-linear least-square regression with the "nls" function from the stats base R package. For parameter *C*, representing basal fluorescence output at 0 inducer concentration, empirical value was used.

$$Fss = (\beta x^n / (K^n + x^n)) + C \quad (1)$$

Where x is Ara (mM) or aTc (nM) inducer concentration.

Toggle and growth assay

Overnight culture grown in absence of inducer was used to inoculate media in 96-well plates supplemented with Ara, aTc and no inducer condition. To toggle, cells were harvested by centrifugation at 4000 RPM for 20 minutes at room temperature and supernatant removed before resuspending in 200 µL LB media, this washing step was repeated for a total of two washes. After final resuspension, 1 µL of washed cells were inoculated to 199 µL fresh media supplemented with the opposite respective inducer. Growth difference metric $\Delta\mu$ was calculated using equation 2.

$$\Delta\mu = \left(\mu_{condition_1} - (\mu_{condition_2} / \mu_{condition_1})\right) * 100 \quad (2)$$

Where μ is max specific growth rate and "condition_1" and "condition_2" denotes sample condition in terms of genotype and induction state. In R, max rates of OD_{600} and normalized fluorescence curves were estimated based on a rolling regression method using the "all_easylinear" function from the growthrates (v.0.8.4, https://CRAN.R-project.org/package=growthrates) R package. Lag times and curve plateaus of OD_{600} and normalized fluorescence curves were determined using the "all_growthmodels" function, fitting the Gompertz growth model[65] with additional lag (λ) parameter.

Pangenome analysis

Genomes of *Stutzerimonas* hosts was downloaded from NCBI. Comparative pangenomic analysis was performed in the Anvi'o (v7.1) environment. pS5 plasmid nucleotide sequence was manually added to each *Stutzerimonas* genome file. For complete data on binned gene clusters obtained comparative pangenomic analysis, see Data and Code Availability. Unless specified, all Anvi'o commands were run using default settings. Briefly, genomes were converted to contigs databases using "anvi-gen-genomes-storage" command, which uses Prodigal[66] to make gene calls. Gene calls were annotated with the Clusters of Orthologous Genes 2020[67] database and KEGG Orthology[35] database. Pangenome analysis was done using "anvi-pan-genome" command with --mcl-inflation parameter set to 10 for high cluster granularity and DIAMOND was run with "-sensitive" flag as recommended when comparing genomes from closely related organisms. Gene clusters were binned according to their number of occurrences across the six genomes in the interactive Anvi'o pangenome display command "anvi-display-pan", from "Bin 1" to "Bin 6". We define gene clusters as belonging to the core genome if it has hits in all six genomes (Bin 6). All other gene clusters were defined as accessory, with gene clusters with only one hit ("Bin 1") further distinguished as unique.

Total RNA extraction and RNA sequencing

Cultures for total RNA harvesting was initiated as described for toggle assay with induction scheme 0.75 mM Ara and 20 nM aTc. Cells were harvested at late exponential growth phase by centrifugation at 10 000 rpm for 1 minute before discarding the supernatant and immediately freezing in liquid nitrogen. Total RNA was extracted using the Quick-RNA Miniprep Kit (Zymo Research, R1055) following manufacturer's instructions. The kit includes a cell lysis step and 15 min on-column DNase I treatment at room temperature. RNA samples in triplicates for each group were sent to Eurofins Genomics (INVIEW Transcriptome Bacteria) for quality control, rRNA depletion (NEBNext® rRNA Depletion Kit (Bacteria), New England Biolabs), cDNA library preparation and sequencing (Illumina NovaSeq 6000 S4 PE150 XP). Across all samples, an average of 23.7 ± 3.1 million reads were obtained.

RNA-Seq and differential gene expression analysis

Trimming, mapping and read counting was done in QIAGEN CLC Genomic Workbench (v22) using the "RNA Seq-Analysis" tool run in default settings (length and similarity fraction = 0.8) to obtain gene expression table files with mapped counts. Annotated reference gnomes were made by annotating genome FASTA files using GFF3 files retrieved from Anvi'o contigs databases using the "anvi-get-sequences-for-gene-calls" command with "--export-gff3" flag. Trimmed reads were mapped to annotated reference genomes. When two reads are equally likely to be mapped to two or more positions, the CLC pipeline randomly maps the read to one of the candidate positions and labels the mapped read as "non-unique". An average of 10.7 ± 1.3 million reads were mapped per sample, of which 9.9 ± 1.6 million reads (93 %) were uniquely mapped. In R (v4.3.1), pangenome gene cluster metadata was mapped to gene expression table by matching Prodigal gene call IDs (see function "r2_RNA_merge_pan_and_count"). Differential gene expression analysis was performed using DESeq2 (v1.40.2) with default settings using pooled counts with a *P*-value threshold of 0.05 and Benjamini and Hochberg adjustment (default to DESeq2) to correct for multiple testing. To allow cross-species comparison of DEG profiles, counts mapped to gene calls within the same gene cluster was pooled for each species.

Statistical analysis

Principal Component Analysis and Procrustes Superimposition analysis was done using the Vegan (v.2.6-4) package in R. The m^2 statistic from PS analysis (scale and symmetric true) was tested for significance by a permutation approach (n = 719, maximum number of iterations). Briefly, observations in one matrix are randomly reordered while maintaining the covariance structure within the matrix and a test statistic is calculated and recorded enough times to obtain a sizeable null distribution. A *P*-value for each statistic is then calculated, representing the probability of obtaining a statistic with a value equal to or more extreme of the experimental value.

DATA AND CODE AVAILABILITY

Experimental data files and R MarkDown scripts used for analysis and plotting are publicly available online on the Open Science Framework database as part of the project name *Chan.Stutz.Pangenome.Chassis* (https://osf.io/yx43n/). Genome and bacterial strain accession numbers can be found in Supplementary Materials. Accession for RNA sequencing data will be made available in the European Nucleotide Archive once ready.

FUNDING

Funding statement: This work was supported by ABSORB – Arctic Carbon Storage from Biomes, which is a strategic funding from UiT – The Arctic University of Norway (https://site.uit.no/absorb/). The computations were performed on resources provided by Sigma2 - the National Infrastructure for High-Performance Computing and Data Storage in Norway.

CONFLICTS OF INTEREST DECLARATION

"The authors declare no competing interests."

ACKNOWLEDGEMENTS

We thank Geoff S. Baldwin from the Imperial College of London for their donation of BASIC parts. We thank the SEVA repository for their donation of pSEVA231 plasmid.

AUTHOR CONTRIBUTIONS

D.T.C.C designed and conducted experiments and wrote the manuscript. H.B designed experiments and wrote the manuscript.

REFERENCES

1. McBride, C. D. & Del Vecchio, D. Predicting Composition of Genetic Circuits with Resource Competition: Demand and Sensitivity. *ACS Synth. Biol.* **10**, 3330–3342 (2021).
2. Adams, B. L. The Next Generation of Synthetic Biology Chassis: Moving Synthetic Biology from the Laboratory to the Field. *ACS Synth. Biol.* **5**, 1328–1330 (2016).
3. Khan, N., Yeung, E., Farris, Y., Fansler, S. J. & Bernstein, H. C. A broad-host-range event detector: expanding and quantifying performance between Escherichia coli and Pseudomonas species. *Synthetic Biology* **5**, ysaa002 (2020).
4. Ke, J. & Yoshikuni, Y. Multi-chassis engineering for heterologous production of microbial natural products. *Current Opinion in Biotechnology* **62**, 88–97 (2020).
5. Calero, P. & Nikel, P. I. Chasing bacterial *chassis* for metabolic engineering: a perspective review from classical to non-traditional microorganisms. *Microb. Biotechnol.* **12**, 98–124 (2019).
6. Abram, K. Z. & Udaondo, Z. Towards a better metabolic engineering reference: the microbial chassis. *Microbial Biotechnology* **13**, 17–18 (2020).
7. Del Vecchio, D. Modularity, context-dependence, and insulation in engineered biological circuits. *Trends in Biotechnology* **33**, 111–119 (2015).
8. Tas, H., Grozinger, L., Stoof, R., de Lorenzo, V. & Goñi-Moreno, Á. Contextual dependencies expand the re-usability of genetic inverters. *Nat Commun* **12**, 355 (2021).
9. Cardinale, S. & Arkin, A. P. Contextualizing context for synthetic biology--identifying causes of failure of synthetic biological systems. *Biotechnol J* **7**, 856–866 (2012).
10. Fatma, Z., Schultz, J. C. & Zhao, H. Recent advances in domesticating non-model microorganisms. *Biotechnology Progress* **36**, e3008 (2020).
11. Russell, J. J. *et al.* Non-model model organisms. *BMC Biology* **15**, 55 (2017).
12. Lammens, E.-M., Nikel, P. I. & Lavigne, R. Exploring the synthetic biology potential of bacteriophages for engineering non-model bacteria. *Nat Commun* **11**, 5294 (2020).

13. Zong, Y. *et al*. Insulated transcriptional elements enable precise design of genetic circuits. *Nat Commun* **8**, 52 (2017).
14. Chan, D. T. C., Baldwin, G. S. & Bernstein, H. C. Revealing the Host-Dependent Nature of an Engineered Genetic Inverter in Concordance with Physiology. *BioDesign Research* **5**, 0016 (2023).
15. Rudra, B. & Gupta, R. S. Phylogenomic and comparative genomic analyses of species of the family Pseudomonadaceae: Proposals for the genera Halopseudomonas gen. nov. and Atopomonas gen. nov., merger of the genus Oblitimonas with the genus Thiopseudomonas, and transfer of some misclassified species of the genus Pseudomonas into other genera. *International Journal of Systematic and Evolutionary Microbiology* **71**, 005011 (2021).
16. Lalucat, J., Gomila, M., Mulet, M., Zaruma, A. & García-Valdés, E. Past, present and future of the boundaries of the Pseudomonas genus: Proposal of Stutzerimonas gen. Nov. *Systematic and Applied Microbiology* **45**, 126289 (2022).
17. Li, X. *et al*. Comparative Genomics of Pseudomonas stutzeri Complex: Taxonomic Assignments and Genetic Diversity. *Frontiers in Microbiology* **12**, (2022).
18. Gomila, M., Mulet, M., García-Valdés, E. & Lalucat, J. Genome-Based Taxonomy of the Genus Stutzerimonas and Proposal of S. frequens sp. nov. and S. degradans sp. nov. and Emended Descriptions of S. perfectomarina and S. chloritidismutans. *Microorganisms* **10**, 1363 (2022).
19. Carlson, C. A., Pierson, L. S., Rosen, J. J. & Ingraham, J. L. Pseudomonas stutzeri and related species undergo natural transformation. *Journal of Bacteriology* **153**, 93–99 (1983).
20. Meier, P., Berndt, C., Weger, N. & Wackernagel, W. Natural transformation of Pseudomonas stutzeri by single-stranded DNA requires type IV pili, competence state and comA. *FEMS Microbiology Letters* **207**, 75–80 (2002).
21. Liu, T., Wang, B., Liu, M., Jiang, K. & Wang, L. Stutzerimonas frequens strain TF18 with superior heterotrophic nitrification-aerobic denitrification ability for the treatment of aquaculture effluent. *Process Biochemistry* **130**, 156–165 (2023).

22. Mulet, M. *et al.* Stutzerimonas decontaminans sp. nov. isolated from marine polluted sediments. *Systematic and Applied Microbiology* **46**, 126400 (2023).

23. Kumari, B. & Chandra, R. Benzo[a]pyrene degradation from hydrocarbon-contaminated soil and their degrading metabolites by Stutzerimonas stutzeri (LOBP-19A). *Waste Management Bulletin* **1**, 115–127 (2023).

24. Brink, D. P., Larsson, E. M., García-Hidalgo, J., Helgadóttir, H. & Gorwa-Grauslund, M. F. Draft Genome Assembly of Stutzerimonas sp. Strain S1 and Achromobacter spanius Strain S4, Two Syringol-Metabolizing Bacteria Isolated from Compost Soil. *Microbiol Resour Announc* **12**, e01150-22.

25. Ge, F. *et al.* Characterization and genomic analysis of Stutzerimonas stutzeri phage vB_PstS_ZQG1, representing a novel viral genus. *Virus Res* **336**, 199226 (2023).

26. Sikorski, J., Lalucat, J. & Wackernagel, W. Genomovars 11 to 18 of Pseudomonas stutzeri, identified among isolates from soil and marine sediment. *International Journal of Systematic and Evolutionary Microbiology* **55**, 1767–1770 (2005).

27. Yu, H. *et al.* Complete Genome Sequence of the Nitrogen-Fixing and Rhizosphere-Associated Bacterium Pseudomonas stutzeri Strain DSM4166 ▿ . *J Bacteriol* **193**, 3422–3423 (2011).

28. Yu, G. *et al.* Systematic identification of endogenous strong constitutive promoters from the diazotrophic rhizosphere bacterium Pseudomonas stutzeri DSM4166 to improve its nitrogenase activity. *Microbial Cell Factories* **22**, 91 (2023).

29. Eren, A. M. *et al.* Community-led, integrated, reproducible multi-omics with anvi'o. *Nat Microbiol* **6**, 3–6 (2021).

30. Jonkheer, E. M. *et al.* The Pectobacterium pangenome, with a focus on Pectobacterium brasiliense, shows a robust core and extensive exchange of genes from a shared gene pool. *BMC Genomics* **22**, 265 (2021).

31. Zhong, C. *et al.* Pan-genome analyses of 24 Shewanella strains re-emphasize the diversification of their functions yet evolutionary dynamics of metal-reducing pathway. *Biotechnol Biofuels* **11**, 193 (2018).
32. Guan, Y. *et al.* Mitigating Host Burden of Genetic Circuits by Engineering Autonegatively Regulated Parts and Improving Functional Prediction. *ACS Synth. Biol.* **11**, 2361–2371 (2022).
33. Lisboa, F. J. G. *et al.* Much beyond Mantel: Bringing Procrustes Association Metric to the Plant and Soil Ecologist's Toolbox. *PLOS ONE* **9**, e101238 (2014).
34. Peres-Neto, P. & Jackson, D. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia* **129**, 169–178 (2001).
35. Aramaki, T. *et al.* KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).
36. Li, Y. *et al.* Cytochrome cd1 Nitrite Reductase NirS Is Involved in Anaerobic Magnetite Biomineralization in Magnetospirillum gryphiswaldense and Requires NirN for Proper d1 Heme Assembly. *J Bacteriol* **195**, 4297–4309 (2013).
37. Bricio, C. *et al.* A Third Subunit in Ancestral Cytochrome c-Dependent Nitric Oxide Reductases. *Appl Environ Microbiol* **80**, 4871–4878 (2014).
38. Silva, L. C. F. *et al.* Physicochemical characterization of Pseudomonas stutzeri UFV5 and analysis of its transcriptome under heterotrophic nitrification/aerobic denitrification pathway induction condition. *Sci Rep* **10**, 2215 (2020).
39. Fu, W. *et al.* Isolation and Identification of an Efficient Aerobic Denitrifying Pseudomonas stutzeri Strain and Characterization of Its Nitrite Degradation. *Catalysts* **11**, 1214 (2021).
40. Jiang, M., Feng, L., Zheng, X. & Chen, Y. Bio-denitrification performance enhanced by graphene-facilitated iron acquisition. *Water Research* **180**, 115916 (2020).
41. Sulavik, M. C. *et al.* Antibiotic Susceptibility Profiles ofEscherichia coli Strains Lacking Multidrug Efflux Pump Genes. *Antimicrobial Agents and Chemotherapy* **45**, 1126–1136 (2001).

42. Fetar, H. *et al.* mexEF-oprN Multidrug Efflux Operon of Pseudomonas aeruginosa: Regulation by the MexT Activator in Response to Nitrosative Stress and Chloramphenicol. *Antimicrobial Agents and Chemotherapy* **55**, 508–514 (2011).

43. Li, M. *et al.* Crystal structure of the transcriptional regulator AcrR from Escherichia coli. *J Mol Biol* **374**, 591–603 (2007).

44. Soumare, A. *et al.* Exploiting Biological Nitrogen Fixation: A Route Towards a Sustainable Agriculture. *Plants (Basel)* **9**, 1011 (2020).

45. Mao, D., Okada, B. K., Wu, Y., Xu, F. & Seyedsayamdost, M. R. Recent Advances in Activating Silent Biosynthetic Gene Clusters in Bacteria. *Curr Opin Microbiol* **45**, 156–163 (2018).

46. Gorochowski, T. E., Avcilar-Kucukgoze, I., Bovenberg, R. A. L., Roubos, J. A. & Ignatova, Z. A Minimal Model of Ribosome Allocation Dynamics Captures Trade-offs in Expression between Endogenous and Synthetic Genes. *ACS Synth. Biol.* **5**, 710–720 (2016).

47. Gyorgy, A. *et al.* Isocost Lines Describe the Cellular Economy of Genetic Circuits. *Biophysical Journal* **109**, 639–646 (2015).

48. Qian, Y., Huang, H.-H., Jiménez, J. I. & Del Vecchio, D. Resource Competition Shapes the Response of Genetic Circuits. *ACS Synth Biol* **6**, 1263–1272 (2017).

49. Carneiro Fidélis Silva, L. *et al.* Heterotrophic nitrifying/aerobic denitrifying bacteria: Ammonium removal under different physical-chemical conditions and molecular characterization. *Journal of Environmental Management* **248**, 109294 (2019).

50. Silva, L. C. F. *et al.* Physicochemical characterization of Pseudomonas stutzeri UFV5 and analysis of its transcriptome under heterotrophic nitrification/aerobic denitrification pathway induction condition. *Sci Rep* **10**, 2215 (2020).

51. Martikainen, P. J. Heterotrophic nitrification – An eternal mystery in the nitrogen cycle. *Soil Biology and Biochemistry* **168**, 108611 (2022).

52. Liu, T., Wang, B., Liu, M., Jiang, K. & Wang, L. Stutzerimonas frequens strain TF18 with superior heterotrophic nitrification-aerobic denitrification ability for the treatment of aquaculture effluent. *Process Biochemistry* **130**, 156–165 (2023).
53. Chen, C. *et al.* Pseudomonas stutzeri GF2 augmented the denitrification of low carbon to nitrogen ratio: Possibility for sewage wastewater treatment. *Bioresource Technology* **333**, 125169 (2021).
54. Hameri, T., Fengos, G. & Hatzimanikatis, V. The effects of model complexity and size on metabolic flux distribution and control: case study in Escherichia coli. *BMC Bioinformatics* **22**, 134 (2021).
55. Lawson, C. E. *et al.* Machine learning for metabolic engineering: A review. *Metabolic Engineering* **63**, 34–60 (2021).
56. Lopatkin, A. J. & Collins, J. J. Predictive biology: modelling, understanding and harnessing microbial complexity. *Nat Rev Microbiol* **18**, 507–520 (2020).
57. Carbonell, P., Radivojevic, T. & García Martín, H. Opportunities at the Intersection of Synthetic Biology, Machine Learning, and Automation. *ACS Synth. Biol.* **8**, 1474–1477 (2019).
58. Ploessl, D., Zhao, Y. & Shao, Z. Engineering of non-model eukaryotes for bioenergy and biochemical production. *Current Opinion in Biotechnology* **79**, 102869 (2023).
59. Riley, L. A. & Guss, A. M. Approaches to genetic tool development for rapid domestication of non-model microorganisms. *Biotechnology for Biofuels* **14**, 30 (2021).
60. Fatma, Z., Schultz, J. C. & Zhao, H. Recent advances in domesticating non-model microorganisms. *Biotechnology Progress* **36**, e3008 (2020).
61. Girard, L. *et al.* Reliable Identification of Environmental Pseudomonas Isolates Using the rpoD Gene. *Microorganisms* **8**, 1166 (2020).
62. Sambrook, J. & Russell, D. W. The inoue method for preparation and transformation of competent e. Coli: ‘ultra-competent’ cells. *CSH Protoc* **2006**, pdb.prot3944 (2006).
63. Storch, M. *et al.* BASIC: A New Biopart Assembly Standard for Idempotent Cloning Provides Accurate, Single-Tier DNA Assembly for Synthetic Biology. *ACS Synth Biol* **4**, 781–787 (2015).

64. Storch, M., Haines, M. C. & Baldwin, G. S. DNA-BOT: A low-cost, automated DNA assembly platform for synthetic biology. *bioRxiv* 832139 (2019) doi:10.1101/832139.

65. Zwietering, M. H., Jongenburger, I., Rombouts, F. M. & van ’t Riet, K. Modeling of the Bacterial Growth Curve. *Appl Environ Microbiol* **56**, 1875–1881 (1990).

66. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).

67. Galperin, M. Y. *et al.* COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res* **49**, D274–D281 (2020).