AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems

Biotechnology | Research Article

Pangenomic landscapes shape performances of a synthetic genetic circuit across *Stutzerimonas* species

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

AUTHOR AFFILIATIONS See affiliation list on p. 20.

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” or "context-dependency". A better understanding of how genomic and physiological contexts underpin the chassis-effect will improve biodesign strategies across diverse microorganisms. Here, we combined a pangenomic-based gene expression analysis with quantitative measurements of performance from an engineered genetic inverter device to uncover how genome structure and function relate 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 the engineered device. Specifically, differential expression of the core genome, gene clusters shared between all hosts, was found to be the main source of significant concordance to the observed differential 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 between hosts in response to the genetic device. Our results show that the chassis-effect can be traced along differences among the most conserved genome-encoded functions 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. As more novel microbes are domesticated for use as biotechnology platforms, synthetic biologists are urged to explore the chassis-design space to optimize their systems and deliver on the promises of synthetic biology. The consequences of the chassis-effect will therefore 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.

KEYWORDS synthetic biology, genetic inverter, transcriptomics, chassis-effect, biodesign, procrustean superimposition, context dependence, non-model organism, host-circuit interactions, broad-host-range

Editor Ryan McClure, Pacific Northwest National Laboratory, Richland, Washington, USA

Address correspondence to Hans C. Bernstein, hans.c.bernstein@uit.no.

The authors declare no conflict of interest.

See the funding table on p. 20.

Received 28 June 2024
Accepted 18 July 2024
Published 21 August 2024

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 the subdiscipline of broad-host-range synthetic biology, which aims to create versatile genetic systems that can operate across diverse organisms and is particularly beneficial toward 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-dependency" (7, 8). The chassis-effect lowers 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 toward 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 moves the current engineering dogma toward the notion that the host itself can be considered a part of tuning circuit functions (7, 8), representing a paradigm shift toward a new understanding of different microbial species and their unique cellular environment as customized hardware for biodesign (13).

Previous efforts to address this knowledge gap have shown that device performances between bacterial hosts can be better explained by the differences in the 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 investigations into the gene expression responses toward the activity of heterologous genetic circuits are needed to uncover insight into which cellular processes underpin chassis-effects. Designing such a study is difficult, as it requires a suit of microbial hosts capable of operating an identical genetic device and producing an observable chassis-effect under the same growth conditions while also sharing significant enough levels 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 genera (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 have 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 a sufficient amount of genomic identity to form a sizeable core (i.e., orthologous genes shared between all hosts) and 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 this group of 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 the 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 a significant correlation between the 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 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, P_{BAD} and P_{Tet}, with each promoter regulating the expression of the other's cognate transcription factor as well as a fluorescent reporter protein in a bicistronic manner. In all, 15 *Stutzerimonas* hosts were screened for plasmid transformability and inverter operability, of which 10 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 pgs16* 24a13 (*S. pgs16*) (26), *Stutzerimonas pgs17* 24a75 (*S. pgs17*) (26), and *Stutzerimonas stutzeri* DSM 4166 (*S. stutzeri*) (27, 28). Phylogenomic analysis of the six *Stutzerimonas* species reveals them to be closely related (Fig. 1b), but comparing pair-wise average nucleotide identity reveals none to be of high similarity to cross the threshold (95%–97%) (29, 30) to be considered of the same species (Fig. S1). Gomila et al.'s work (2022) (16, 18) pioneered the clarification of the *Stutzerimonas* clade and also reported that our six hosts can each be considered distinct phylogenomic species. The names of our *Stutzerimonas* hosts follow the proposed names assigned by (18).

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 (31), leading to a total of 25,344 gene calls functionally annotated using the 2020 clusters of orthologous genes (COGs) database (Fig. 1c). These gene calls were grouped into 6,469 gene clusters or "pangenomic orthologous groups." A gene cluster is grouped into a "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. 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 make up 31.1% (2,013) of total gene clusters. The clustering of the six hosts by their presence/absence of the identified orthologous gene clusters differs by the phylogenomic clustering of GToTree (Fig. 1d). The number of host-specific gene calls ranged from 3,731 in *S. degradans* to 4,457 in *S. pgs16*, with all hosts sharing on average 67.2% ± 4.4% of genes. 23% of all gene clusters were not assigned to any COG category (assigned to the "NA" group) (Fig. 1e), with an additional 3.6% and 4.9% assigned to COG category S (Function Unknown) and R (General Prediction Only), respectively. Of these three groups of lesser-known gene clusters, 81% belong to the accessory genome, with approximately half 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

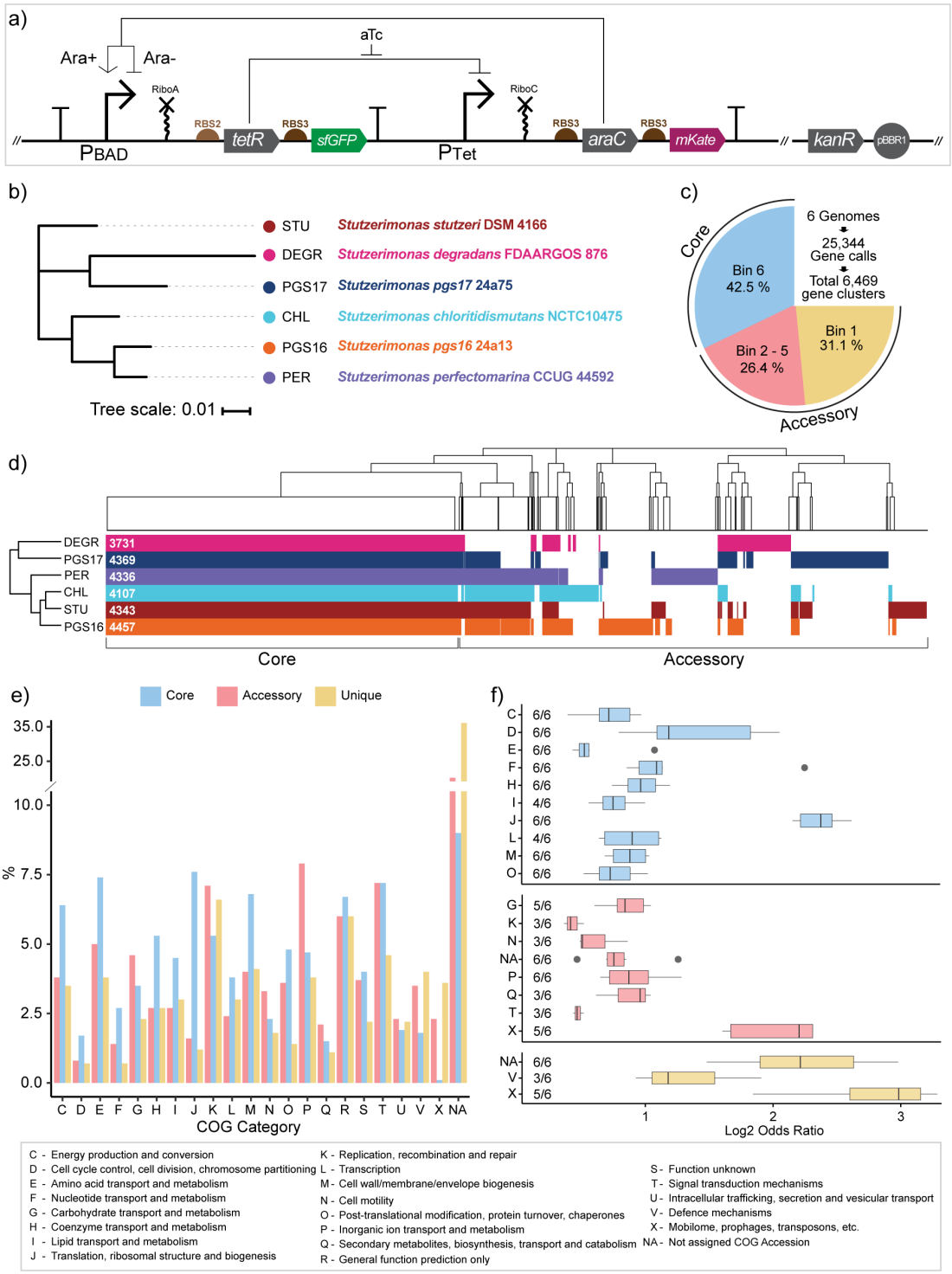


FIG 1 The genetic inverter and pangenome of selected *Stutzerimonas* hosts. (a) Schematic representation of Ara-aTc genetic inverter design. In the presence of Ara (Ara+), Ara-bound AraC upregulates its cognate promoter (P_{BAD}), leading to sfGFP and TetR expression and, in turn, creates a distinct measurable fluorescent state and leads to the downregulation of the P_{Tet} promoter. In the 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 the upregulation of one leads to the downregulation of the other. (b) Inferred phylogenomic tree of the six *Stutzerimonas* hosts. The scale bar is in units of the number of amino acid substitutions per site between two sequences. (c) Composition of core, accessory, and unique gene clusters from pangenome analysis (Anvi'o) of our six *Stutzerimonas* species. Bin 6 means all six hosts contribute with at least one gene call to the gene cluster and Bin 5 means any combination of five hosts contribute with at least one gene call and so on. Bin 5 to Bin 1 are grouped as the accessory genome, with gene clusters belonging to Bin 1 further distinguished as unique. Percentages indicate the portion of

(Continued on next page)

FIG 1 (Continued)

the 6,469 gene clusters assigned to the three frequency groups (core, accessory, and unique). (d) Clustered presence/absence matrix of the 6,469 orthologous gene clusters with columns representing gene clusters. The number indicates the number of gene calls for each host. (e) Percentage composition of cluster of orthologous genes (COG) categories of core, accessory, and unique group. (f) Enrichment analysis of COG categories within each frequency group by Fisher exact test. "*n*/6" indicates the number of hosts in which COG category was found significantly (*P*-value < 0.05, Bonferroni correction) enriched within the group. Only COG categories enriched in three or more hosts are shown. Gray points represent outliers. COG category description is provided at the bottom of the figure. CHL = *Stutzerimonas chloritidismutans* NCTC10475; PER = *Stutzerimonas perfectomarina* CCUG 44592; DEGR = *Stutzerimonas degradans* FDAARGOS 876; PGS16 = *Stutzerimonas pgs16* 24a13, PGS17 = *Stutzerimonas pgs17* 24a75; STU = *Stutzerimonas stutzeri* DSM 4166.

house genes that confer a specific advantage to the organism within its niche environment. These gene clusters might therefore only be expressed under certain conditions, making them difficult to study 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. 1f). The Fisher exact test (*P*-value > 0.05, with Bonferroni Correction) revealed that within the core genome, COG categories consistently enriched (i.e., significantly enriched in ≥3 species) for 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. We note that phenotypic characterization done by (18) reveals that *S. chloritidismutans*, *S. degradans,* and *S. perfectomarina* tested negative in the metabolization of arabinose, with no characterization data available for the remaining three. Comparative BLAST search using the sequence of *E. coli araABCD* operon as a query against the genomes of all six *Stutzerimonas* hosts yielded no significant similarities in sequence as well. Few COG categories were found to be consistently overrepresented within the unique genome, likely due to the majority of genes being of unknown function, but the few are category X (Mobilome: Prophages and Transposons), which hints toward past viral infection events, and V (Defense Mechanisms). Other prokaryotic pangenome studies report similar patterns of enriched COG categories in their defined shared and accessory genomes (32–34). The distinct functional identity of genomic sub-groups gives merit to the hypothesis that the 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

Comparing the quantified performance of the engineered genetic inverter operated by each host under a standardized environment revealed a clear chassis-effect. The performance of the inverter operating within *Escherichia coli* DH5α (*E. coli*) was also quantified as a reference. Induction response dynamics were characterized by fitting the Hill function $[(\beta x^n / (K^n + x^n)) + C]$ to normalized induction curves (Fig. 2a through d). Parameter 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 (activation coefficient), 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. We observe markedly different performance profile from the genetic inverter depending on host context — that is, a strong chassis-effect. For instance, the observed K_{aTc} values range from 1.46 to 35.11 nM aTc among the *Stutzerimonas* hosts (Fig. 2a and b), suggesting host-specific factors affecting

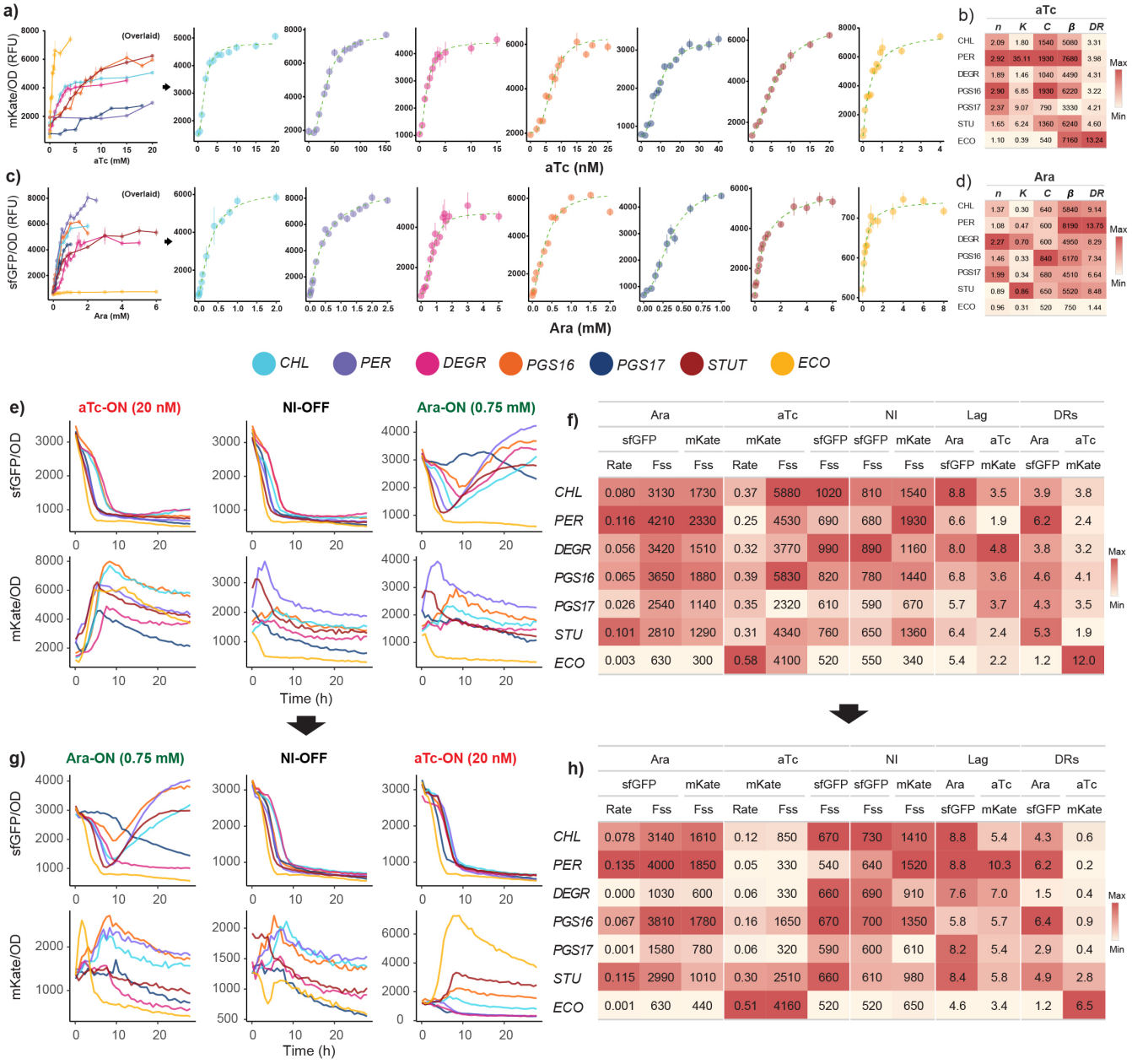


FIG 2 The chassis-effect is observed through the measurable performance of the genetic inverter between closely related *Stutzerimonas* hosts. (a) aTc induction curves, the left-most plot shows all induction curves overlaid up to a given inducer concentration. All induction curves to the right show individual curves with scaled axes. 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. DR = dynamic range; NI = no induction; Fss = late phase steady-state fluorescence; Rate = max specific rate; DRs = specific dynamic range.

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 0.39. The K_{Ara} metrics were relatively more uniform across hosts, with only an overall 2.8-fold difference between the smallest and largest K_{Ara} values (Fig. 2c and d). We define an additional metric, DR_{Ara} and DR_{aTc} (dynamic range) as the ratio of the estimated β and empirical C expressed as a fold-change value, representing the largest possible fold-change difference measurable

from respective reporter proteins. At saturated inducer concentrations, the highest β_{Ara} achieved among the *Stutzerimonas* hosts was *S. perfectomarina* at 8,190 RFU, which also achieves the highest DR_{Ara} value of 13,75. Overall, *E. coli* had the lowest β_{Ara} at 750 RFU. Meanwhile the range of observed DR_{aTc} values were relatively uniform, ranging from 3.22 to 4.60 among the *Stutzerimonas*.

We next performed a toggle assay to demonstrate the invertibility of the device and to further quantify the chassis-effect (Fig. 2e through h). Initial OFF cells (grown in the 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 were estimated across induction states. These metrics (Fig. 2f) include the maximum steady-state fluorescence at the late growth phase (F_{ss}), maximum rate of fluorescence (Rate) at the exponential phase, specific dynamic range (DR_S), and lag time (Lag). Each of these metrics has 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—that is, the host-specific background fluorescence/expression leakage. Meanwhile, fluorescence output from cells grown under the presence of an antagonistic inducer (e.g., sfGFP output in the presence of aTc) indicates deficient transcriptional control. 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 media containing respective opposite inducers (Fig. 2g and 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* 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 to 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 were performed (Fig. S2), and for *S. chloritidismutans*, *S. degradans,* and *S. stutzeri*, the mKate output attenuation was relieved under induction schemes with decreased Ara concentration. However, the other three hosts experienced consistent attenuation across toggling schemes. As 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 the 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 through d). The reference host *E. coli* showed the highest growth on LB media compared to all *Stutzerimonas* hosts. The addition of BB23 backbone led to decreased specific growth rates of all strains compared to their respective wild-type counterparts, with *S. pgs17* exhibiting the most reduced growth rate (Fig. 3e and f). The consistent growth reduction is expected due to the added burden of maintaining the vector backbone in the presence of kanamycin. *E. coli* and *S. chloritidismutans* exhibit the greatest growth burden upon the addition of the inverter device into the backbone (complete pS5 plasmid) under NI conditions, likely due to a larger plasmid payload (35) and leakage expression of the device. However, some hosts experience close to zero additional growth burden, and while *S. chloritidismutans* had a relatively high amount of leakage expression (Fig. 2), *E. coli* also had the lowest leakage expression, suggesting the mechanism in which the backbone vector and device impose growth inhibition varies between hosts. Induction of the genetic inverter exacerbated the growth inhibition (Fig. 3e and f). For instance, the growth rate of *S. pgs17*

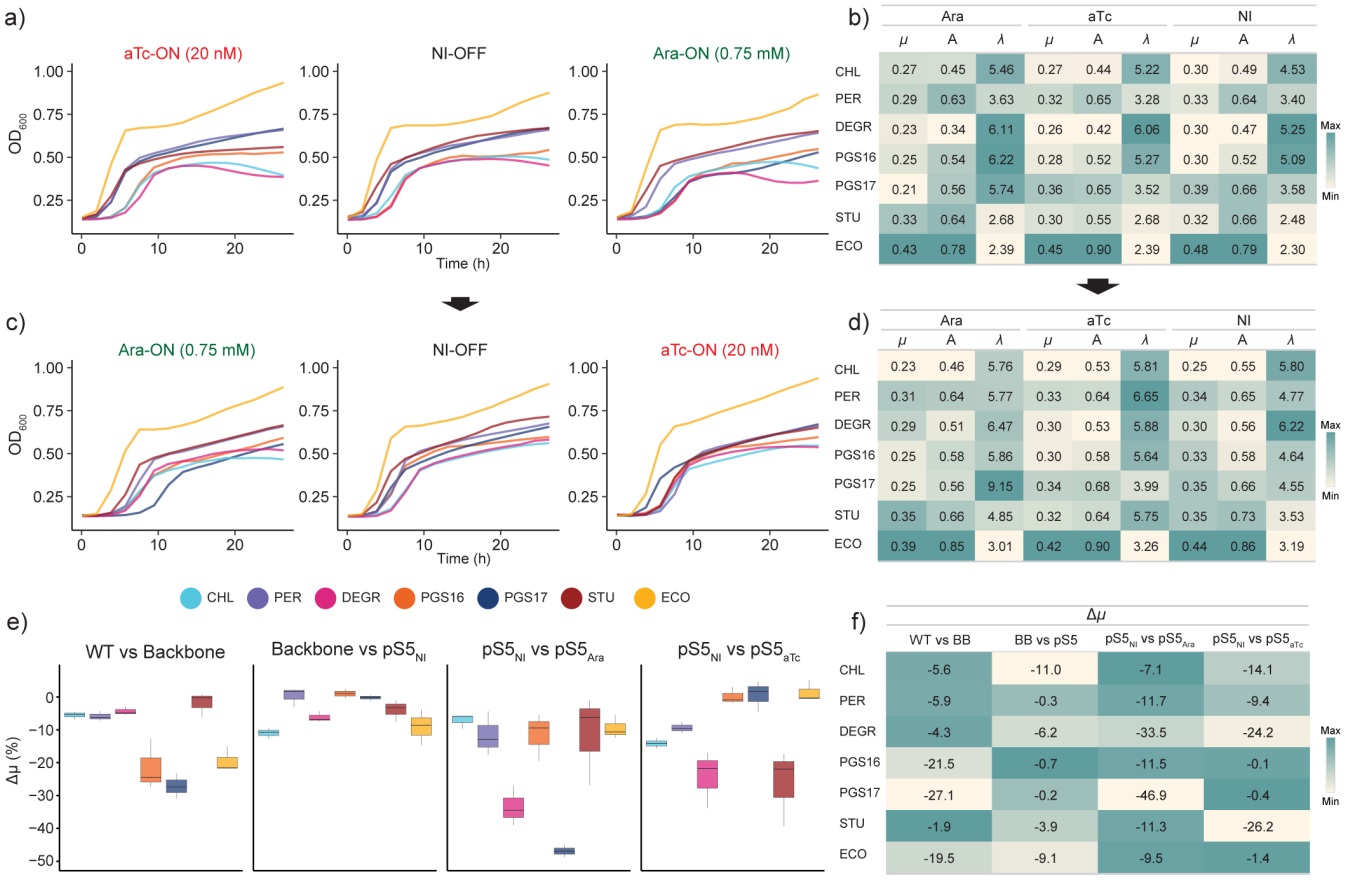


FIG 3 The growth dynamics are uniquely affected as a result of the host-specific operation of the genetic inverter. (a) Growth curves of initial OFF cells diluted to respective induction states. Hosts are color-coded. (b) Estimated growth metrics for each host from growth curves in (a). The color scale is relative to each column, as for all subsequent subpanels. (c) Growth curves of toggled cells diluted to respective opposite inducer to toggle induction state and (d) corresponding estimated growth metrics. (e) Boxplots of growth difference between host genotypes and/or induction state captured by the Δμ metric, defined as a relative percentage change in growth rate. Whiskers indicate minimum and maximum values. (f) Table overview of metrics from (e). WT = wild type, BB23 = pBBR1, and KanR cloning vector. $pS5_{NI}$ = pS5 plasmid in the absence of an inducer. $pS5_{Ara}$ = pS5 plasmid in the presence of Ara (0.75 mM), $pS5_{aTc}$ = pS5 plasmid in the presence of aTc (20 nM). NI = no induction; μ = max specific growth rate; A = carrying capacity; λ = lag time.

was almost halved when induced with Ara when compared with the NI condition but showed little response when treated with aTc. Overall, these results exemplify that the genetic inverter's mode of operation uniquely affects the growth of each host.

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 have on the observed host-specific inverter performance.

Unique transcriptional patterns arise from both the 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 the 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 programmed operability (Fig. 4). Cells 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 (Fig. S3). 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. pgs17*) 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 high degree of transcriptional readthrough in aTc-induced cells when transcribing from the P_{Tet} promoter, highlighting the importance of context insulation (36). A relatively large difference in TPM values was also observed between polycistronic repressor-reporter pairs (Fig. S3).

The mRNA abundance of genes encoded by the genetic inverter intuitively plays 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 global transcriptome response of all hosts to

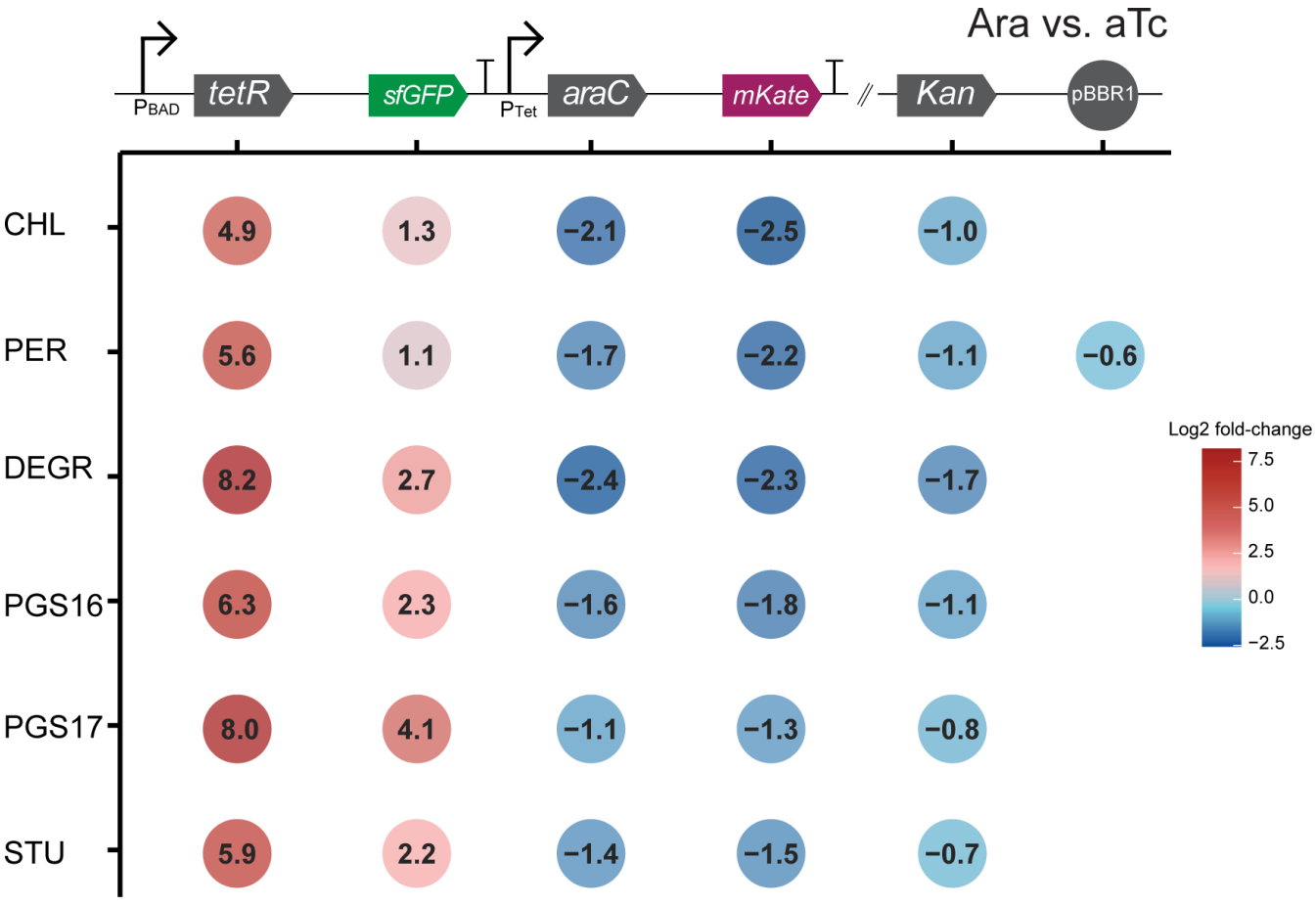


FIG 4 The chassis-effect is measurable in the differential response of genes designed into the engineered genetic inverter. Log2 fold-change values of the six genes encoded in the pS5 plasmid between hosts comparing Ara against aTc-induced cells. An empty position indicates a non-significant differential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

better understand how the 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. 5). Differential expression analysis with an adjusted *P*-value > 0.05 threshold leads to a total of 4,825 significantly DEGs distributed across hosts, with 3,995 DEGs belonging to the core group distributed among 1,672 gene clusters (Fig. 5a and b). 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. The number of DEGs is also unevenly distributed between the hosts, varying from 589 (*S. chloritidismutans*) to 1,106 (*S. pgs17*) 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 we therefore include responses to inducers as a contributor to the chassis-effect. A different clustering pattern is observed upon performing hierarchical clustering of the hosts in terms of their differential expression of the core and accessory group (Fig. 5c and d) except for *S. stutzeri*, whose expression response clustered uniquely into its own branch from the other five hosts in both core and accessory groups. Overall, this result further solidifies the difference in expression response among the hosts.

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 show a uniform distribution across the 22 COG categories while a more varied response pattern in the accessory genome is observed. Among core DEGs, 45.3% ± 1.5 % of the DEGs can be found in categories C, E, J, M, and 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 differences in the composition of the upregulated and downregulated 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 because of inverter activity. A larger variation in the distribution of DEGs was observed within a COG category as well. For instance, 31.3% of DEGs in category K occurred in *S. perfectomarina* alone, exemplifying the diversity in gene expression response patterns. 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 belongs to the major carbon, nitrogen, and energy metabolism, cell division, and housekeeping genes, which can more intuitively underpin the chassis-effect, given that the majority of DEGs stem from the core genome.

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

Enrichment analysis and Procrustes Superimposition (PS) analysis both indicate that the key genome-encoded functions responsible for the chassis-effect originate from differences in the expression of the core genes shared among all hosts. PS analysis is a statistical test that can be used to determine the strength of correlation between two multivariate data sets by comparing the goodness-of-fit between two configurations

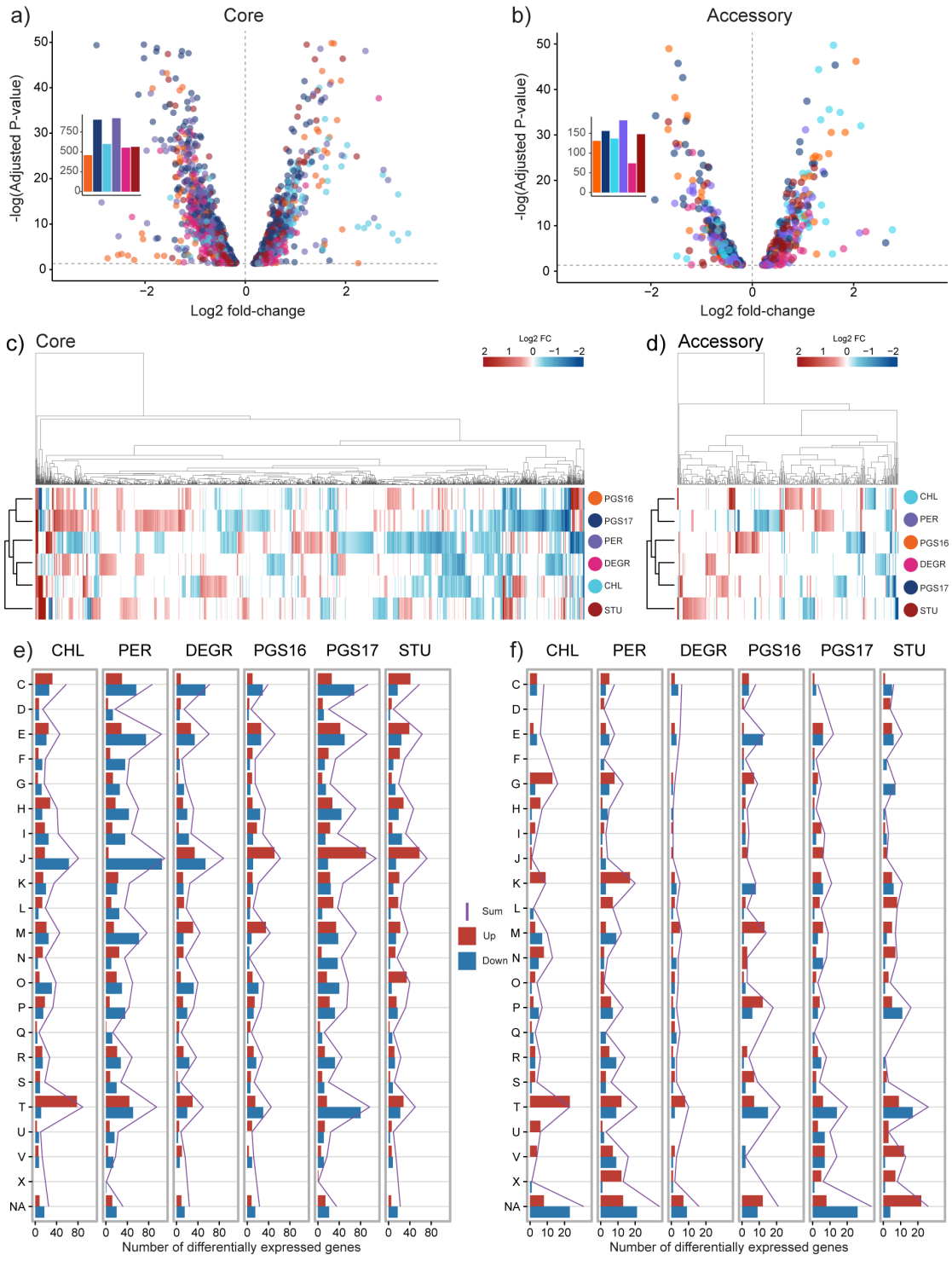


FIG 5 Global differential gene expression analysis of Ara against aTc-induced cells reveals diverse transcriptional profiles as a result of inverter operation. Volcano plots visualizing log2 fold-change distribution of significant DEGs in (a) core genome and (b) accessory genome. Inset bar charts show the number of differentially expressed genes for each host. Clustered heatmap of log2 fold-change values of (c) core and (d) accessory gene clusters significantly expressed by at least one host (*P*-value < 0.05, Benjamini and Hochberg adjusted). White bars indicates non-significant expression in the host. For the accessory genome, white bars indicates either non-significance or the gene cluster has no hits for that host. The distribution of DEGs across COG categories and hosts is 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 the number of DEGs.

consisting of n items (our hosts) on a coordinate plane. When the distances between points in the configuration carry information of (dis)similarity, such as points plotted along principal components, PS analysis can be used to determine whether the (dis)similarities between two datasets correlate (37, 38). PS analysis outputs the Gower statistic (m^2), which is the sum of squared vector residuals that remains after optimal fitting and can be tested for significance through a permutation method that maintains the internal co-variance of each data set.

Fisher's exact test showed that core gene clusters were significantly enriched within the significantly DEGs for all hosts, while accessory and unique gene clusters were significantly underrepresented (Fig. 6a through c). The strongest source of gene expression response to the inverter device operation thereby stems from core genes, suggesting the core genome to be significant in driving the chassis-effect. Building upon this result, we conducted PS analysis to answer the question of whether the differential inverter device performance between our hosts correlates with the differences between them in terms of their gene expression response (i.e., do hosts with more similar gene expression response also have more similar performance?).

Principal component analysis (PCA) was first performed to project the captured differences among our hosts along the first two principal components for each data set. These configurations in ordinate space were then compared by PS. The process of fit optimization is illustrated in Fig. 6d. To avoid artificial inflation of distance, the unique gene clusters were omitted from the data set of accessory DEGs, as these gene clusters inherently have no comparison between hosts. A significant correlation is observed when comparing differences in inverter performance and differences in expression response for all DEGs (Fig. 6d; *P*-value = 0.043, m^2 = 0.328). 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 = 0.035, m^2 = 0.312) (Fig. 6e and f). 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 patterns of their shared core genome also have significantly more similar performances. Hence, for 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. In addition, PS analysis applied to the captured growth metrics shows a significant correlation with inverter device performance (Fig. S4), corroborating our previous result that observed differences in the physiological state of hosts can be used to potentially predict device performance (14).

We also 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 PS analysis. Ten of the 23 COG groups were found to be significant (Fig. S5). Category E (Amino acid Transport and Metabolism) is especially intuitive because the main carbon and energy source provided by 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

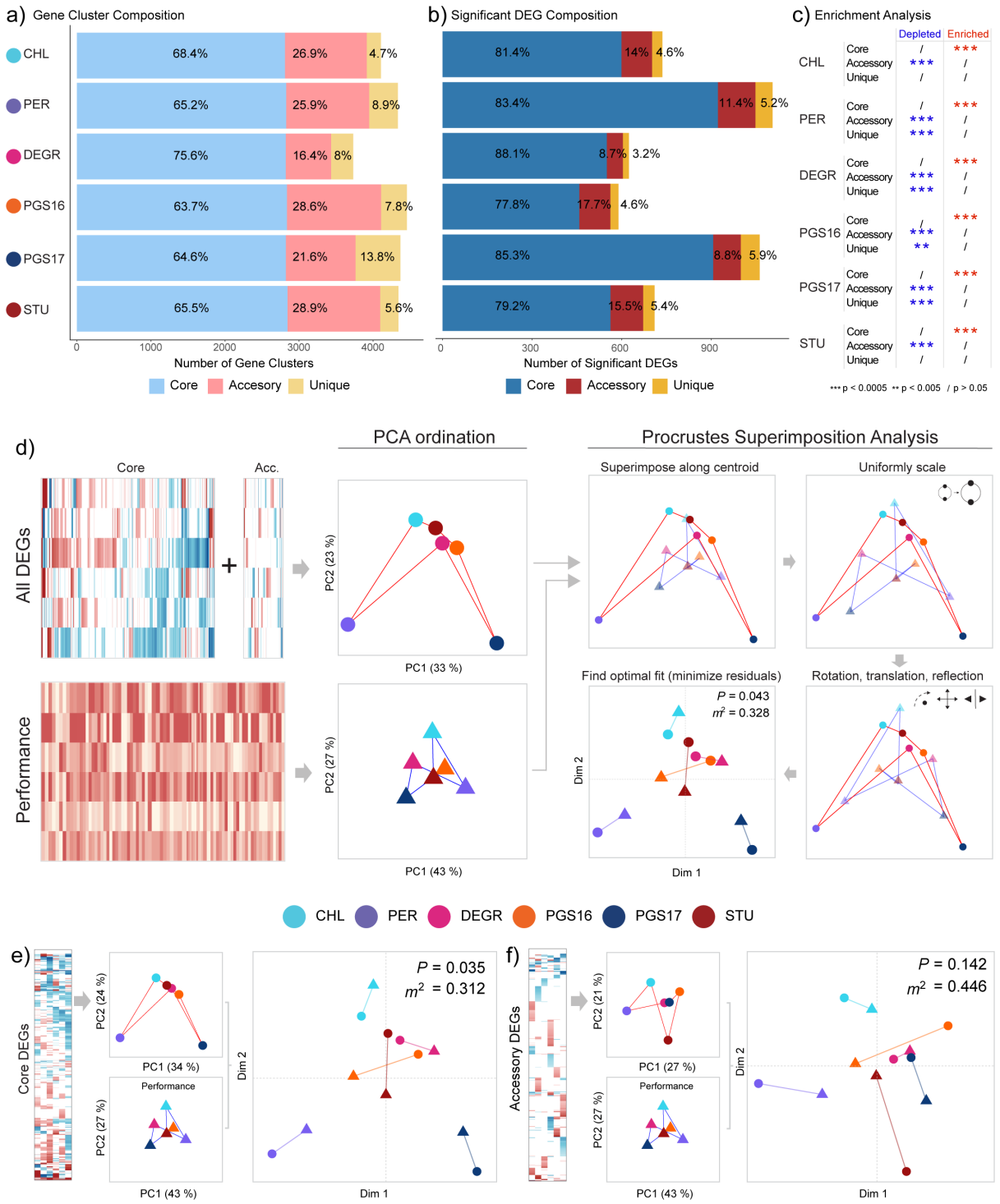


FIG 6 Procrustes analysis reveals significant concordance between similarity in inverter performance and similarity in core genome response between hosts. (a) Composition of each host's genome in terms of core, accessory, and unique gene clusters. (b) Composition of significantly differentially expressed gene clusters in terms of core, accessory, and unique gene clusters. (c) Enrichment analysis results by Fisher exact test, testing for depletion (underrepresentation) and enrichment (overrepresentation) for the three frequency groups for each host. (d) Procrustes superimposition analysis comparing hosts in terms of all significant differential gene expression responses against inverter device performance metrics. For the accessory genome, unique gene clusters were omitted to reduce artificial inflation of distance. The key steps in PS analysis are schematically illustrated. Performance metric data set and differential gene expression data set 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 to minimize the sum of squared vector residuals (the m^2 statistic) between each respective point (host). The significance of the obtained

(Continued on next page)

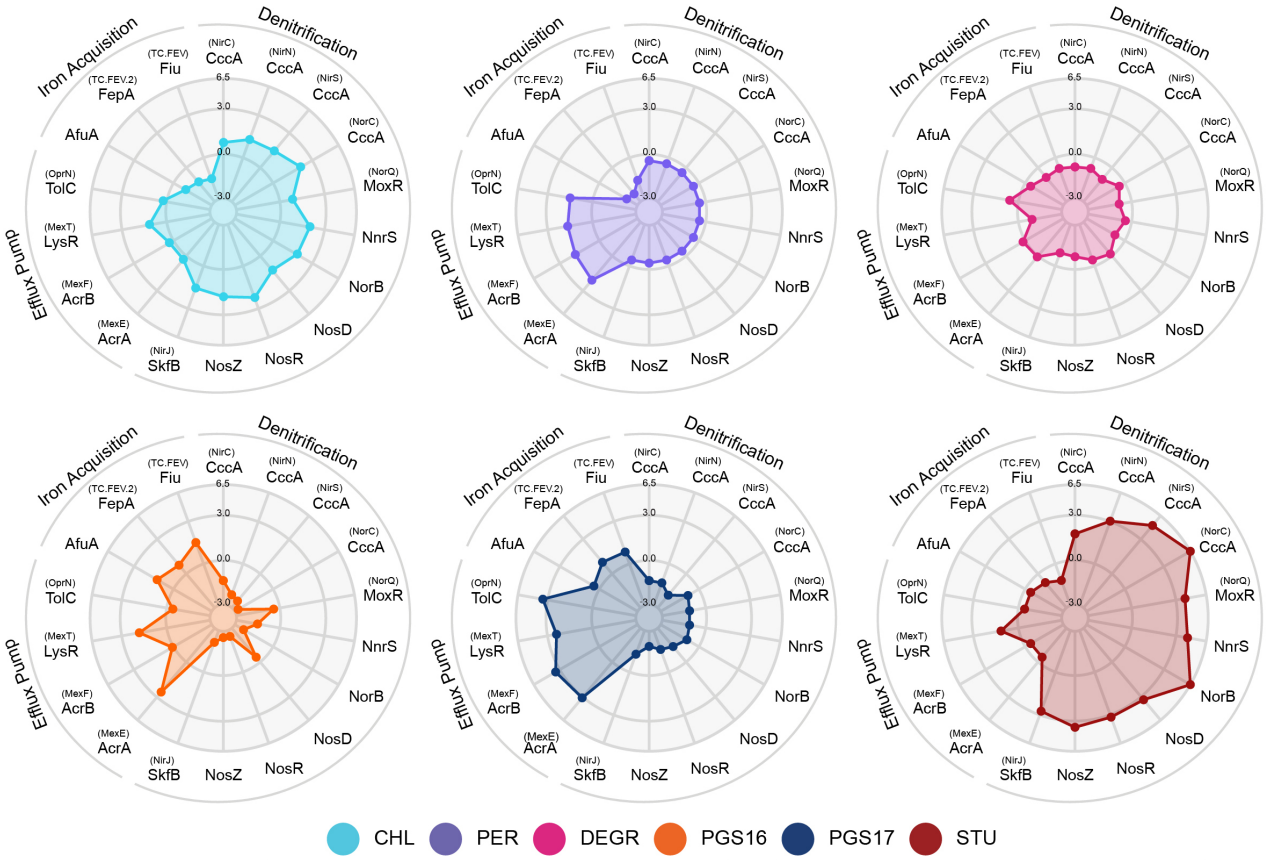
FIG 6 (Continued)

statistic is determined through a permutation method. Colored lines between points have been added to visualize an arbitrary "configuration" formed by each data set, which connects each point in the following arbitrary order "-CHL-PER-DEGR-PGS16-PGS17-STU-." PS analysis comparing inverter performance against significantly differentially expressed (e) core and (f) accessory gene clusters. *P* = *P*-value, m^2 = *Gower statistic*.

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 (39) 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 GC_00001064, GC_00001456, GC_00001627, and GC_00001727 describes all these gene clusters to encode for cytochrome c protein (CccA). Meanwhile, the corresponding KO annotations for the four *CccA*-annotated gene clusters are cytochrome c55X (NirC, K19344), dihydroheme 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 to the cytochrome c protein family (40, 41). 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 discussed here all match their respective database records, indicating high degree of agreement between the two annotation methods.

Genes involved in denitrification were upregulated in *S. chloritidismutans* and *S. stutzeri* and downregulated in *S. pgs16* and *S. pgs17* but showed little change in *S. perfectomarina* and *S. degradans* when comparing Ara against aTc induction states. This result was not anticipated and exemplifies that the 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, 42, 43). 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. perfectomarina* and ranked low. Genes involved in iron acquisition showed the opposite trend, being upregulated in *S. pgs16*and *S. pgs17* and downregulated to some degree in all other hosts. Denitrification activity requires bioavailable iron for heme biosynthesis (44); 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_00000078, GC_00000115, and GC_00000605. 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 (45, 46). Another high-ranking gene cluster was GC_00002718, annotated by COG as a transcriptional regulator part of the LysR family, but as the MexT protein by KO. Incidentally, MexT is a LysR-type transcriptional regulator regulating the MexEF-OprN operon, indicating that gene clusters GC_00000078, GC_00000115, and GC_00000605 indeed encode for a MexEF-OprN efflux pump, as AcrAB-TolC is regulated by the transcription factor AcrR, which instead belongs to the TetR family of transcriptional regulators (47). Interestingly, hosts that exhibit downregulated or low change in the expression of denitrification gene clusters were also the hosts with the highest upregulation of the AcrAB-TolC/MexEF-OprN encoding gene clusters (hosts *S. perfectomarina*, *S. pgs16,* and *S. pgs17*). Fetar et al. (46) have also reported that nitrosative stress in the form of nitric oxide accumulation is a direct inducer of MexEF-oprN expression (46), suggesting that hosts expressing the efflux pump could be a response to nitric oxide produced as part of denitrification activity.



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

FIG 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.

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 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 as a biotechnology host with its known applications in

the bioremediation of contaminants (18). *S. pgs17* is inferred to possess the *nifDKH* genes according to KEGG pathway mapping (48) and therefore has potential application as a biological nitrogen fixation agent to replace agrochemical similar to *S. stutzeri*, the latter having already received much attention for its nitrogen fixation capability (27).

Given the standardized growth conditions on rich LB medium, 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 the 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 that may uniquely influence how each host responds to a change in substrate availability or changing growth conditions (49). 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 cultured under identical conditions 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 have demonstrated how resource dynamics such as ribosome abundance (50) can affect genetic circuit performance and even parameterized resource competition to build predictive models (51, 52).

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 of 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 (Fig. S6), suggesting a lack of nitrification capability. However, there have been numerous reports of other *Stutzerimonas* (reported in the defunct name *Pseudomonas stutzeri*) capable of converting ammonia to nitrogen gas through heterotrophic nitrification and aerobic denitrification (HNAD) despite their genomes encoding neither *amo* nor *hao* (21, 42, 53, 54). 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 is a pragmatic phenotype due to its bioremediation potential (55), 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 (56), 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 denitrification and efflux pumps but given that it is specific to the host context, 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 (57, 58). 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 (59).

We have here 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 (10, 60, 61), heralding the advancement toward 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 to optimize their engineered systems. As the field of biodesign progresses toward 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

An overview of the species used in this study can be found in Table S2. The six *Stutzerimonas* hosts selected for further study were verified by their *rpoD* sequence, determined through sanger sequencing (62) (Table S3). Cells were cultured in Lysogeny-Broth (LB) at 35°C unless specified otherwise. BB23 backbone and pS5-carrying strains were cultivated in the 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 (1,096 cpm, 1 mm) at 9 mm read height. Working stock solutions of 1 M L-Arabinose (VWR, A11921) and 1 mM aTc (VWR, CAYM10009542) were prepared by dissolving in MilliQ water and 70% ethanol, respectively. Cloning was performed using *E. coli* DH5α, made chemically competent, and transformed following the Inoue method (63). *Stutzerimonas* species were transformed *via* the 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 pSEVA231 as a template with primers B_SEVA_F and B_SEVA_R. Plasmid pS5 was assembled in the Biopart Assembly

Standard for Idempotent Cloning (BASIC) (64, 65) environment as previously described (14). Sequence and accession of pS5 components can be found in Table S5.

Induction assays

Overnight culture grown in the absence of an inducer was used to inoculate the media with various concentrations of aTc and Ara in 96-well plates. The normalized steady-state fluorescence at the late growth phase (Fss) averaged over a time window of 6–12 hours was used as a response variable of induction curves. In R (v4.3.1), Hill coefficient (*n*), activation coefficient (*K*), and max steady-state fluorescence output (*β*) were estimated by fitting the Hill function (1) using non-linear least-square regression with the "nls" function from the base R *stats* package. For parameter *C*, representing basal fluorescence output at 0 inducer concentration, an empirical value was used.

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

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

Toggle and growth assay

Overnight culture grown in the 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 4,000 RPM for 20 minutes at room temperature and the supernatant was 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 was inoculated to 199 µL fresh media supplemented with the opposite respective inducer.

The growth difference metric $\Delta\mu$ was calculated using equation 2.

$$\Delta\mu = \left(\mu_{condition\_1} - \left(\mu_{condition\_2} / \mu_{condition\_1}\right)\right) * 100 \quad (2)$$

Where μ is max specific growth rate and "condition_1" and "condition_2" denote two different sample conditions describing 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) 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 (66) with an additional lag (λ) parameter.

Phylogenomic tree and pangenome

Genomes of *Stutzerimonas* hosts were downloaded from NCBI. GToTree (67) was used to infer phylogenomic relationship using the 174 single-copy gene set (Gammaproteobacteria Hidden Markov Model set) under default settings, which infers tree through Maximum-Likelihood using FastTree (68). 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 from 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 the "anvi-gen-genomes-storage" command, which uses Prodigal (69) to make gene calls. Gene calls were annotated with the Clusters of Orthologous Genes 2020 (70) database through Anvi'o and the KEGG Orthology (39) database through KofamKOALA. Pangenome analysis was done using "anvi-pan-genome" command with the --mcl-inflation parameter set to 10 for high cluster granularity and DIAMOND was run with the "-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 accessories, with gene clusters with only one hit ("Bin 1") further distinguished as unique.

Total RNA extraction and RNA sequencing

Cultures for total RNA harvesting were initiated as described for toggle assay with induction scheme 0.75 mM Ara and 20 nM aTc to ensure transcriptome profile representative of the toggle assay. Cells were harvested at the late exponential growth phase to ensure enough cell mass 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 the manufacturer's instructions. The kit includes a cell lysis step and 15-minute on-column DNase I treatment at room temperature. RNA samples in biological 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 was obtained.

RNA-Seq and differential gene expression analysis

Trimming, mapping, and read counting were 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 genomes 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 the "--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 was mapped per sample, of which 9.9 ± 1.6 million reads (93%) were uniquely mapped. In R, pangenome gene cluster metadata were mapped to the 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 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 were pooled for each species.

Data and statistical analysis

Hierarchical clustering was done using the "hclust" function from the base R *stats* package, using Euclidean distance by the "complete" method. Fisher exact test was done using the "fisher.test" function from the *stats* package, testing for both depletion and enrichment. Principal component analysis and Procrustes Superimposition analysis were done using the *Vegan* (v.2.6.4) package. The first two principal components for each data set were used for downstream analysis. The m^2 statistic from PS analysis (scale and symmetric set 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.

ACKNOWLEDGMENTS

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.

Funding statement: This work was supported by ABSORB—Arctic Carbon Storage from Biomes, which is 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.

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

AUTHOR AFFILIATIONS

[1]Faculty of Biosciences, Fisheries and Economics, UiT - The Arctic University of Norway, Tromsø, Norway
[2]The Arctic Centre for Sustainable Energy, UiT - The Arctic University of Norway, Tromsø, Norway

AUTHOR ORCIDs

Dennis Tin Chat Chan http://orcid.org/0000-0002-0841-5898
Hans C. Bernstein http://orcid.org/0000-0003-2913-7708

FUNDING

Funder	Grant(s)	Author(s)
ABSORB - Arctic carbon storage from biomes		Hans C. Bernstein

DATA 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. RNA sequencing data can be accessed through project PRJEB74560 (stu-RNAseq) from the European Nucleotide Archive.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Supplemental figures (mSystems00849-24-s0001_v2.docx). Fig. S1 to S6.
Supplemental tables (mSystems00849-24-s0002_v2.xlsx). Tables S1 to S5.

REFERENCES

1. McBride CD, Del Vecchio D. 2021. Predicting composition of genetic circuits with resource competition: demand and sensitivity. ACS Synth Biol 10:3330–3342. https://doi.org/10.1021/acssynbio.1c00281
2. Adams BL. 2016. The next generation of synthetic biology chassis: moving synthetic biology from the laboratory to the field. ACS Synth Biol 5:1328–1330. https://doi.org/10.1021/acssynbio.6b00256
3. Khan N, Yeung E, Farris Y, Fansler SJ, Bernstein HC. 2020. A broad-host-range event detector: expanding and quantifying performance between *Escherichia coli* and *Pseudomonas* species. Synth Biol 5:ysaa002. https://doi.org/10.1093/synbio/ysaa002
4. Ke J, Yoshikuni Y. 2020. Multi-chassis engineering for heterologous production of microbial natural products. Curr Opin Biotechnol 62:88–97. https://doi.org/10.1016/j.copbio.2019.09.005
5. Calero P, Nikel PI. 2019. Chasing bacterial chassis for metabolic engineering: a perspective review from classical to non-traditional microorganisms. Microb Biotechnol 12:98–124. https://doi.org/10.1111/1751-7915.13292
6. Abram KZ, Udaondo Z. 2020. Towards a better metabolic engineering reference: the microbial chassis. Microb Biotechnol 13:17–18. https://doi.org/10.1111/1751-7915.13363
7. Del Vecchio D. 2015. Modularity, context-dependence, and insulation in engineered biological circuits. Trends Biotechnol 33:111–119. https://doi.org/10.1016/j.tibtech.2014.11.009
8. Tas H, Grozinger L, Stoof R, de Lorenzo V, Goñi-Moreno Á. 2021. Contextual dependencies expand the re-usability of genetic inverters. Nat Commun 12:355. https://doi.org/10.1038/s41467-020-20656-5

9. Cardinale S, Arkin AP. 2012. Contextualizing context for synthetic biology--identifying causes of failure of synthetic biological systems. Biotechnol J 7:856–866. https://doi.org/10.1002/biot.201200085
10. Fatma Z, Schultz JC, Zhao H. 2020. Recent advances in domesticating non-model microorganisms. Biotechnol Prog 36:e3008. https://doi.org/10.1002/btpr.3008
11. Russell JJ, Theriot JA, Sood P, Marshall WF, Landweber LF, Fritz-Laylin L, Polka JK, Oliferenko S, Gerbich T, Gladfelter A, Umen J, Bezanilla M, Lancaster MA, He S, Gibson MC, Goldstein B, Tanaka EM, Hu C-K, Brunet A. 2017. Non-model model organisms. BMC Biol 15:55. https://doi.org/10.1186/s12915-017-0391-5
12. Lammens E-M, Nikel PI, Lavigne R. 2020. Exploring the synthetic biology potential of bacteriophages for engineering non-model bacteria. Nat Commun 11:5294. https://doi.org/10.1038/s41467-020-19124-x
13. Zong Y, Zhang HM, Lyu C, Ji X, Hou J, Guo X, Ouyang Q, Lou C. 2017. Insulated transcriptional elements enable precise design of genetic circuits. Nat Commun 8:52. https://doi.org/10.1038/s41467-017-00063-z
14. Chan DTC, Baldwin GS, Bernstein HC. 2023. Revealing the host-dependent nature of an engineered genetic inverter in concordance with physiology. Biodes Res 5:0016. https://doi.org/10.34133/bdr.0016
15. Rudra B, Gupta RS. 2021. 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. Int J Syst Evol Microbiol 71:005011. https://doi.org/10.1099/ijsem.0.005011
16. Lalucat J, Gomila M, Mulet M, Zaruma A, García-Valdés E. 2022. Past, present and future of the boundaries of the *Pseudomonas genus*: proposal of *Stutzerimonas* gen. nov. Syst Appl Microbiol 45:126289. https://doi.org/10.1016/j.syapm.2021.126289
17. Li X, Yang Z, Wang Z, Li W, Zhang G, Yan H. 2021. Comparative genomics of *Pseudomonas stutzeri* complex: taxonomic assignments and genetic diversity. Front Microbiol 12:755874. https://doi.org/10.3389/fmicb.2021.755874
18. Gomila M, Mulet M, García-Valdés E, Lalucat J. 2022. 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. https://doi.org/10.3390/microorganisms10071363
19. Carlson CA, Pierson LS, Rosen JJ, Ingraham JL. 1983. *Pseudomonas stutzeri* and related species undergo natural transformation. J Bacteriol 153:93–99. https://doi.org/10.1128/jb.153.1.93-99.1983
20. Meier P, Berndt C, Weger N, Wackernagel W. 2002. Natural transformation of *Pseudomonas stutzeri* by single-stranded DNA requires type IV pili, competence state and *comA*. FEMS Microbiol Lett 207:75–80. https://doi.org/10.1111/j.1574-6968.2002.tb11031.x
21. Liu T, Wang B, Liu M, Jiang K, Wang L. 2023. *Stutzerimonas frequens* strain TF18 with superior heterotrophic nitrification-aerobic denitrification ability for the treatment of aquaculture effluent. Proc Biochem 130:156–165. https://doi.org/10.1016/j.procbio.2023.04.018
22. Mulet M, Gomila M, Lalucat J, Bosch R, Rossello-Mora R, García-Valdés E. 2023. *Stutzerimonas decontaminans* sp. nov. isolated from marine polluted sediments. Syst Appl Microbiol 46:126400. https://doi.org/10.1016/j.syapm.2023.126400
23. Kumari B, Chandra R. 2023. Benzo[a]pyrene degradation from hydrocarbon-contaminated soil and their degrading metabolites by *Stutzerimonas stutzeri* (LOBP-19A). Waste Man Bull 1:115–127. https://doi.org/10.1016/j.wmb.2023.07.006
24. Brink DP, Larsson EM, García-Hidalgo J, Helgadóttir H, Gorwa-Grauslund MF. 2023. 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:e0115022. https://doi.org/10.1128/mra.01150-22
25. Ge F, Guo R, Liang Y, Chen Y, Shao H, Sung YY, Mok WJ, Wong LL, McMinn A, Wang M. 2023. Characterization and genomic analysis of *Stutzerimonas stutzeri* phage vB_PstS_ZQG1, representing a novel viral genus. Virus Res 336:199226. https://doi.org/10.1016/j.virusres.2023.199226
26. Sikorski J, Lalucat J, Wackernagel W. 2005. Genomovars 11 to 18 of *Pseudomonas stutzeri*, identified among isolates from soil and marine sediment. Int J Syst Evol Microbiol 55:1767–1770. https://doi.org/10.1099/ijs.0.63535-0
27. Yu H, Yuan M, Lu W, Yang J, Dai S, Li Q, Yang Z, Dong J, Sun L, Deng Z, Zhang W, Chen M, Ping S, Han Y, Zhan Y, Yan Y, Jin Q, Lin M. 2011. Complete genome sequence of the nitrogen-fixing and rhizosphere-associated bacterium *Pseudomonas stutzeri* strain DSM4166. J Bacteriol 193:3422–3423. https://doi.org/10.1128/JB.05039-11
28. Yu G, Li X, Duan Q, Fu J, Zhang Y, Wang H, Luan J. 2023. Systematic identification of endogenous strong constitutive promoters from the diazotrophic rhizosphere bacterium *Pseudomonas stutzeri* DSM4166 to improve its nitrogenase activity. Microb Cell Fact 22:91. https://doi.org/10.1186/s12934-023-02085-3
29. Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A 106:19126–19131. https://doi.org/10.1073/pnas.0906412106
30. Rodriguez-R LM, Conrad RE, Viver T, Feistel DJ, Lindner BG, Venter SN, Orellana LH, Amann R, Rossello-Mora R, Konstantinidis KT. 2024. An ANI gap within bacterial species that advances the definitions of intra-species units. mBio 15:e0269623. https://doi.org/10.1128/mbio.02696-23
31. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M, Fogarty EC, et al. 2021. Community-led, integrated, reproducible multi-omics with anvi'o. Nat Microbiol 6:3–6. https://doi.org/10.1038/s41564-020-00834-3
32. Jonkheer EM, Brankovics B, Houwers IM, van der Wolf JM, Bonants PJM, Vreeburg RAM, Bollema R, de Haan JR, Berke L, Smit S, de Ridder D, van der Lee TAJ. 2021. 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. https://doi.org/10.1186/s12864-021-07583-5
33. Zhong C, Han M, Yu S, Yang P, Li H, Ning K. 2018. 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. https://doi.org/10.1186/s13068-018-1201-1
34. Nikolaidis M, Mossialos D, Oliver SG, Amoutzias GD. 2020. Comparative analysis of the core proteomes among the *Pseudomonas* major evolutionary groups reveals species-specific adaptations for *Pseudomonas aeruginosa* and *Pseudomonas chlororaphis*. Diversity 12:289. https://doi.org/10.3390/d12080289
35. Guan Y, Chen X, Shao B, Ji X, Xiang Y, Jiang G, Xu L, Lin Z, Ouyang Q, Lou C. 2022. Mitigating host burden of genetic circuits by engineering autonegatively regulated parts and improving functional prediction. ACS Synth Biol 11:2361–2371. https://doi.org/10.1021/acssynbio.2c00073
36. Park Y, Espah Borujeni A, Gorochowski TE, Shin J, Voigt CA. 2020. Precision design of stable genetic circuits carried in highly-insulated *E.coli* genomic landing pads. Mol Syst Biol 16:e9584. https://doi.org/10.15252/msb.20209584
37. Lisboa FJG, Peres-Neto PR, Chaer GM, Jesus E da C, Mitchell RJ, Chapman SJ, Berbara RLL. 2014. Much beyond mantel: bringing procrustes association metric to the plant and soil ecologist's toolbox. PLoS One 9:e101238. https://doi.org/10.1371/journal.pone.0101238
38. Peres-Neto PR, Jackson DA. 2001. How well do multivariate data sets match? The advantages of a procrustean superimposition approach over the mantel test. Oecologia 129:169–178. https://doi.org/10.1007/s004420100720
39. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. 2020. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics 36:2251–2252. https://doi.org/10.1093/bioinformatics/btz859
40. Li Y, Bali S, Borg S, Katzmann E, Ferguson SJ, Schüler D. 2013. Cytochrome CD1 nitrite reductase NirS is involved in anaerobic magnetite biomineralization in magnetospirillum gryphiswaldense and requires NirN for proper d_1 heme assembly. J Bacteriol 195:4297–4309. https://doi.org/10.1128/JB.00686-13
41. Bricio C, Alvarez L, San Martin M, Schurig-Briccio LA, Gennis RB, Berenguer J. 2014. A third subunit in ancestral cytochrome *c*-dependent nitric oxide reductases. Appl Environ Microbiol 80:4871–4878. https://doi.org/10.1128/AEM.00790-14
42. Silva LCF, Lima HS, Mendes TA de O, Sartoratto A, Sousa MP, de Souza RS, de Paula SO, de Oliveira VM, Silva CC. 2020. Physicochemical

characterization of *Pseudomonas stutzeri* UFV5 and analysis of its transcriptome under heterotrophic nitrification/aerobic denitrification pathway induction condition. Sci Rep 10:2215. https://doi.org/10.1038/s41598-020-59279-7
43. Fu W, Wang Q, Chen S, Wang Y, Wang Y, Duan P, Yi G, Liu C, Zhang X, Rao Z. 2021. Isolation and identification of an efficient aerobic denitrifying *Pseudomonas stutzeri* strain and characterization of its nitrite degradation. Catalysts 11:1214. https://doi.org/10.3390/catal11101214
44. Jiang M, Feng L, Zheng X, Chen Y. 2020. Bio-denitrification performance enhanced by graphene-facilitated iron acquisition. Water Res 180:115916. https://doi.org/10.1016/j.watres.2020.115916
45. Sulavik MC, Houseweart C, Cramer C, Jiwani N, Murgolo N, Greene J, DiDomenico B, Shaw KJ, Miller GH, Hare R, Shimer G. 2001. Antibiotic susceptibility profiles of *Escherichia coli* strains lacking multidrug efflux pump genes. Antimicrob Agents Chemother 45:1126–1136. https://doi.org/10.1128/AAC.45.4.1126-1136.2001
46. Fetar H, Gilmour C, Klinoski R, Daigle DM, Dean CR, Poole K. 2011. *mexEF-oprN* multidrug efflux operon of *Pseudomonas aeruginosa*: regulation by the MexT activator in response to nitrosative stress and chloramphenicol. Antimicrob Agents Chemother 55:508–514. https://doi.org/10.1128/AAC.00830-10
47. Li M, Gu R, Su C-C, Routh MD, Harris KC, Jewell ES, McDermott G, Yu EW. 2007. Crystal structure of the transcriptional regulator AcrR from *Escherichia coli*. J Mol Biol 374:591–603. https://doi.org/10.1016/j.jmb.2007.09.064
48. Soumare A, Diedhiou AG, Thuita M, Hafidi M, Ouhdouch Y, Gopalakrishnan S, Kouisni L. 2020. Exploiting biological nitrogen fixation: a route towards a sustainable agriculture. Plants (Basel) 9:1011. https://doi.org/10.3390/plants9081011
49. Mao D, Okada BK, Wu Y, Xu F, Seyedsayamdost MR. 2018. Recent advances in activating silent biosynthetic gene clusters in bacteria. Curr Opin Microbiol 45:156–163. https://doi.org/10.1016/j.mib.2018.05.001
50. Gorochowski TE, Avcilar-Kucukgoze I, Bovenberg RAL, Roubos JA, Ignatova Z. 2016. A minimal model of ribosome allocation dynamics captures trade-offs in expression between endogenous and synthetic genes. ACS Synth Biol 5:710–720. https://doi.org/10.1021/acssynbio.6b00040
51. Gyorgy A, Jiménez JI, Yazbek J, Huang H-H, Chung H, Weiss R, Del Vecchio D. 2015. Isocost lines describe the cellular economy of genetic circuits. Biophys J 109:639–646. https://doi.org/10.1016/j.bpj.2015.06.034
52. Qian Y, Huang H-H, Jiménez JI, Del Vecchio D. 2017. Resource competition shapes the response of genetic circuits. ACS Synth Biol 6:1263–1272. https://doi.org/10.1021/acssynbio.6b00361
53. Carneiro Fidélis Silva L, Santiago Lima H, Antônio de Oliveira Mendes T, Sartoratto A, de Paula Sousa M, Suhett de Souza R, Oliveira de Paula S, Maia de Oliveira V, Canêdo da Silva C. 2019. Heterotrophic nitrifying/aerobic denitrifying bacteria: ammonium removal under different physical-chemical conditions and molecular characterization. J Environ Manage 248:109294. https://doi.org/10.1016/j.jenvman.2019.109294
54. Martikainen PJ. 2022. Heterotrophic nitrification – an eternal mystery in the nitrogen cycle. Soil Bio Biochem 168:108611. https://doi.org/10.1016/j.soilbio.2022.108611
55. Chen C, Ali A, Su J, Wang Y, Huang T, Gao J. 2021. *Pseudomonas stutzeri* GF2 augmented the denitrification of low carbon to nitrogen ratio: possibility for sewage wastewater treatment. Biores Techn 333:125169. https://doi.org/10.1016/j.biortech.2021.125169
56. Hameri T, Fengos G, Hatzimanikatis V. 2021. The effects of model complexity and size on metabolic flux distribution and control: case study in *Escherichia coli*. BMC Bioinformatics 22:134. https://doi.org/10.1186/s12859-021-04066-y
57. Lawson CE, Martí JM, Radivojevic T, Jonnalagadda SVR, Gentz R, Hillson NJ, Peisert S, Kim J, Simmons BA, Petzold CJ, Singer SW, Mukhopadhyay A, Tanjore D, Dunn JG, Garcia Martin H. 2021. Machine learning for metabolic engineering: a review. Metab Eng 63:34–60. https://doi.org/10.1016/j.ymben.2020.10.005
58. Lopatkin AJ, Collins JJ. 2020. Predictive biology: modelling, understanding and harnessing microbial complexity. Nat Rev Microbiol 18:507–520. https://doi.org/10.1038/s41579-020-0372-5
59. Carbonell P, Radivojevic T, García Martín H. 2019. Opportunities at the intersection of synthetic biology, machine learning, and automation. ACS Synth Biol 8:1474–1477. https://doi.org/10.1021/acssynbio.8b00540
60. Ploessl D, Zhao Y, Shao Z. 2023. Engineering of non-model eukaryotes for bioenergy and biochemical production. Curr Opin Biotechnol 79:102869. https://doi.org/10.1016/j.copbio.2022.102869
61. Riley LA, Guss AM. 2021. Approaches to genetic tool development for rapid domestication of non-model microorganisms. Biotechnol Biofuels 14:30. https://doi.org/10.1186/s13068-020-01872-z
62. Girard L, Lood C, Rokni-Zadeh H, van Noort V, Lavigne R, De Mot R. 2020. Reliable identification of environmental *Pseudomonas* isolates using the *rpoD* gene. Microorganisms 8:1166. https://doi.org/10.3390/microorganisms8081166
63. Sambrook J, Russell DW. 2006. The inoue method for preparation and transformation of competent *E. coli*: "ultra-competent" cells. Cold Spring Harb Protoc 2006:db. https://doi.org/10.1101/pdb.prot3944
64. Storch M, Casini A, Mackrow B, Fleming T, Trewhitt H, Ellis T, Baldwin GS. 2015. BASIC: a new biopart assembly standard for Idempotent cloning provides accurate, single-tier DNA assembly for synthetic biology. ACS Synth Biol 4:781–787. https://doi.org/10.1021/sb500356d
65. Storch M, Haines MC, Baldwin GS. 2020. DNA-BOT: a low-cost, automated DNA assembly platform for synthetic biology. Synth Biol (Oxf) 5:ysaa010. https://doi.org/10.1093/synbio/ysaa010
66. Zwietering MH, Jongenburger I, Rombouts FM, van 't Riet K. 1990. Modeling of the bacterial growth curve. Appl Environ Microbiol 56:1875–1881. https://doi.org/10.1128/aem.56.6.1875-1881.1990
67. Lee MD. 2019. GToTree: a user-friendly workflow for phylogenomics. Bioinformatics 35:4162–4164. https://doi.org/10.1093/bioinformatics/btz188
68. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490
69. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119
70. Galperin MY, Wolf YI, Makarova KS, VeraAlvarez R, Landsman D, Koonin EV. 2021. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. Nucleic Acids Res 49:D274–D281. https://doi.org/10.1093/nar/gkaa1018