

Population genomics-guided engineering of phenazine biosynthesis in *Pseudomonas chlororaphis*

Sarah Thorwall^{a,1}, Varun Trivedi^{a,1}, Eva Ottum^a, Ian Wheeldon^{a,b,c,*}

^a Department of Chemical and Environmental Engineering, University of California, Riverside, CA 92521, USA

^b Center for Industrial Biotechnology, University of California, Riverside, CA 92521, USA

^c Integrative Institute for Genome Biology, University of California, Riverside, CA 92521, USA

ARTICLE INFO

Keywords:

Bacterial genome assembly

Pangenome

GWAS

Pseudomonas chlororaphis

ABSTRACT

The emergence of next-generation sequencing (NGS) technologies has made it possible to not only sequence entire genomes, but also identify metabolic engineering targets across the pangenome of a microbial population. This study leverages NGS data as well as existing molecular biology and bioinformatics tools to identify and validate genomic signatures for improving phenazine biosynthesis in *Pseudomonas chlororaphis*. We sequenced a diverse collection of 34 *Pseudomonas* isolates using short- and long-read sequencing techniques and assembled whole genomes using the NGS reads. In addition, we assayed three industrially relevant phenotypes (phenazine production, biofilm formation, and growth temperature) for these isolates in two different media conditions. We then provided the whole genomes and phenazine production data to a unitig-based microbial genome-wide association study (mGWAS) tool to identify novel genomic signatures responsible for phenazine production in *P. chlororaphis*. Post-processing of the mGWAS analysis results yielded 330 significant hits influencing the biosynthesis of one or more phenazine compounds. Based on a quantitative metric (called the phenotype score), we elucidated the most influential hits for phenazine production and experimentally validated them *in vivo* in the most optimal phenazine producing strain. Two genes significantly increased phenazine-1-carboxamide (PCN) production: a histidine transporter (ProY_1), and a putative carboxypeptidase (PS_04251). A putative MarR-family transcriptional regulator decreased PCN titer when overexpressed in a high PCN producing isolate. Overall, this work seeks to demonstrate the utility of a population genomics approach as an effective strategy in enabling the identification of targets for metabolic engineering of bioproduction hosts.

1. Introduction

The development of next-generation sequencing and CRISPR genome editing has enabled entire microbial genomes to be sequenced and manipulated, resulting in genome-wide metabolic engineering approaches often within non-traditional hosts. With further advancements in DNA sequencing technologies it is now economically feasible for a single research group to sequence small collections of tens to hundreds of microbial isolates, sometimes even in-house with portable sequencing devices. New metabolic engineering strategies could take advantage of this increasing accessibility of microbial whole-genome sequencing data and existing bioinformatics tools to analyze this data to identify metabolic engineering targets from a collection of genomes in a “pangenome”-wide or population genomics approach.

A first step in the design of a new metabolic engineering project is the selection of an appropriate host that natively exhibits a phenotype of interest. This work seeks to improve phenazine production in the bacterium *Pseudomonas chlororaphis* by identifying non-intuitive genetic targets from a collection of *P. chlororaphis* isolates as part of a population genomics approach to metabolic engineering. Phenazines are redox-active, often colorful, secondary metabolites with applications in agriculture as antifungal agents and potential applications as redox mediators in flow cell batteries and bioelectrochemical devices (Hollas et al., 2018; Rabaey et al., 2005; Clifford et al., 2021). *P. chlororaphis* is a commercially available biocontrol species that would make a good potential phenazine production host as it natively produces multiple phenazine derivatives, is non-pathogenic to humans and plants, can utilize the inexpensive carbon source glycerol, and has available synthetic biology tools for genetic manipulation. Many strains also have

* Corresponding author. Department of Chemical and Environmental Engineering, University of California, Riverside, CA 92521, USA.

E-mail addresses: sthor007@ucr.edu (S. Thorwall), vtriv002@ucr.edu (V. Trivedi), wheeldon@ucr.edu (I. Wheeldon).

¹ These authors contributed equally.

Abbreviations

CDS	coding sequence(s)
KMB	King's media B
KMB + Fe	King's media B supplemented with ferric iron
mGWAS	microbial genome-wide association studies
PCA	phenazine-1-carboxylic acid
PCN	phenazine-1-carboxamide
2-HP	2-hydroxyphenazine
2-HPCA	2-hydroxyphenazine-1-carboxylic acid

traits that are detrimental to industrial bioprocessing, such as biofilm formation and low growth temperatures. Here, we sequenced the genomes of 34 *Pseudomonas* isolates, characterized their bioprocess-relevant phenotypes (phenazine production, biofilm formation, and growth temperature), and conducted microbial genome-wide association studies (mGWAS) to select an optimal host strain for phenazine production and identify genetic manipulations that increase phenazine biosynthesis.

P. chlororaphis has already been successfully engineered for phenazine production, with metabolic engineering works pursuing rational design strategies. Replacing genes within the phenazine biosynthesis operon can modulate final phenazine composition and allow non-native phenazines to be produced, including 1-hydroxyphenazine (Wan et al., 2021) and phenazine-1,6-dicarboxylic acid derivatives iodinin and 1,6-dimethoxyphenazine (Guo et al., 2022)). Regulation of the phenazine biosynthesis operon provides opportunities to improve phenazine production, including *phzR* and *phzI* which directly regulate expression through quorum sensing (Yu et al., 2018) and components of the Gac/Rsm pathway (e.g. *rpeA*, *rsmE*, *lon* protease, *psrA*, *parS*, *gacA* (Wan et al., 2021; K. Liu et al., 2016; Li et al., 2020; W.-H. Liu et al., 2021)) which indirectly interact with PhzR/PhzI in response to other environmental factors. Increasing carbon flux through the shikimate pathway, such as by overexpressing *aroB*, *aroD*, *aroE*, *ppsA*, and *tktA* (K. Liu et al., 2016; Li et al., 2020), also improves phenazine production by increasing flux through phenazine biosynthesis. Collectively, these approaches combined with fermentation optimization have been able to produce grams per liter titers of phenazines, including 0.68 g/L of 2-hydroxyphenazine (W.-H. Liu et al., 2021), 3.6 g/L of 1-hydroxyphenazine (Wan et al., 2021), and 11.45 g/L of PCN in *P. chlororaphis* (Li et al., 2020).

Our population genomics approach uses mGWAS to identify metabolic engineering targets from our genomic and phenotypic data. GWAS correlate genomic and phenotypic datasets to identify causal genetic variants (Lees et al., 2016). While GWAS are most commonly used to identify human disease risk factors, recent bioinformatics tools have been developed to adapt these studies to bacteria (Brynildsrud et al., 2016; Lees et al., 2016; Jaillard et al., 2018). We input our sequenced genomes and phenotypic data into DBGWAS (Jaillard et al., 2018), a unitig-based mGWAS tool, to identify genomic loci that are significantly associated with phenazine production. This approach requires no prior knowledge of relevant biosynthetic pathways and could identify previously unknown targets for metabolic engineering throughout a single genome and the genomes of a population of isolates. We further sought to experimentally validate the top mGWAS hits by overexpressing associated genes and measuring phenazine production with respect to the wildtype control. Genes associated to hits identified in this study present new targets for strain engineering of *P. chlororaphis* to improve phenazine bioproduction.

2. Results and discussion

2.1. Curating a *P. chlororaphis* strain collection

We began our population genomics approach to metabolic engineering by collecting a library of *P. chlororaphis* strains, a known over-producer of phenazines. We purchased all unique strains that were accessible to us from international culture collections, resulting in 26 strains purchased from three culture collections (Table S1). Eight of these strains had multiple colony morphologies present which appeared to vary in pigment production. Because these variants could differ in the phenotypes of interest and consequently may have associated genetic variation, each of these variants was treated as a separate isolate with the strain number indicating the original culture collection strain designation followed by a superscript ¹ and ² arbitrarily assigned to the differing morphologies. 16S rRNA sequencing identified 33 isolates as *P. chlororaphis* and 1 as *P. synxantha* for a total of 34 *Pseudomonas* isolates used in this study.

2.2. Phenotyping for phenazine production, biofilm formation, and growth temperature

For the phenazine biosynthesis phenotyping, we quantify the four phenazine compounds naturally produced by *P. chlororaphis*: 2-hydroxyphenazine (2-HP), 2-hydroxyphenazine-1-carboxylic acid (2-HPCA), phenazine-1-carboxylic acid (PCA) and phenazine-1-carboxamide (PCN) (Fig. 1) (Mavrodi et al. 2006). In phenazine-producing pseudomonads, the core phenazine biosynthesis operon is responsible for synthesizing PCA which serves as the precursor for other phenazine derivatives. *P. chlororaphis* strains typically produce either PCN or a combination of 2-HPCA and 2-HP depending on whether *phzH* or *phzO*, respectively, is present and functional.

We first characterized phenazine production in King's Media B (KMB), the standard culture media for fluorescent pseudomonads (Fig. 2). Under these conditions, fewer than half the isolates produced more than 10 mg/L of phenazines. These low titers suggest that phenotyping in KMB may underestimate the phenazine production capacity of our strain collection. To improve phenazine production and consequently the quality of this dataset for our mGWAS analysis, we supplemented KMB with 100 μ M ferric iron, which has been previously reported to enhance phenazine production in some strains of *P. chlororaphis* (Chin-A-Woeng et al., 1998; van Rij et al., 2004). KMB + Fe media improved phenazine production in 24 isolates, while the remaining 10 isolates produced no significant phenazines even after the addition of iron. Due to its positive effects for most of the strains and its neutral effects on the remaining strains, we did additional phenotyping in KMB + Fe as well as KMB.

Out of all strains we characterized, strain DSM 21509 was found to produce significantly higher titers of PCN (477 ± 163 mg/L; one-way ANOVA followed by Tukey's test, $p < 0.05$) than all other strains in KMB + Fe. DSM 21509 was also one of the strains producing significantly higher PCN titers in KMB, along with NCCB 100368¹ (>95 mg/L for both strains; $p < 0.001$). Strain DSM 21509 was thus deemed as the best host strain overall for PCN production. DSM 21509 is the type strain of *P. chlororaphis* subsp. *piscium*. This strain was isolated from the intestine of a European perch from Lake Neuchâtel, Switzerland, in 2005 (Burr et al., 2010). Strains ATCC 17417, NCCB 88062¹, ATCC 15926 and ATCC 13985 were found to produce significantly higher titers of combined PCA/2-HPCA/2-HP in KMB + Fe (>150 mg/L for each strain; $p < 0.05$) compared to the remaining strains, while strain NCCB 88062¹ alone produced a significantly higher titer of combined PCA/2-HPCA/2-HP in KMB (80 ± 6 mg/L; $p < 0.0001$). Strain NCCB 88062¹ was therefore considered the most optimum strain for

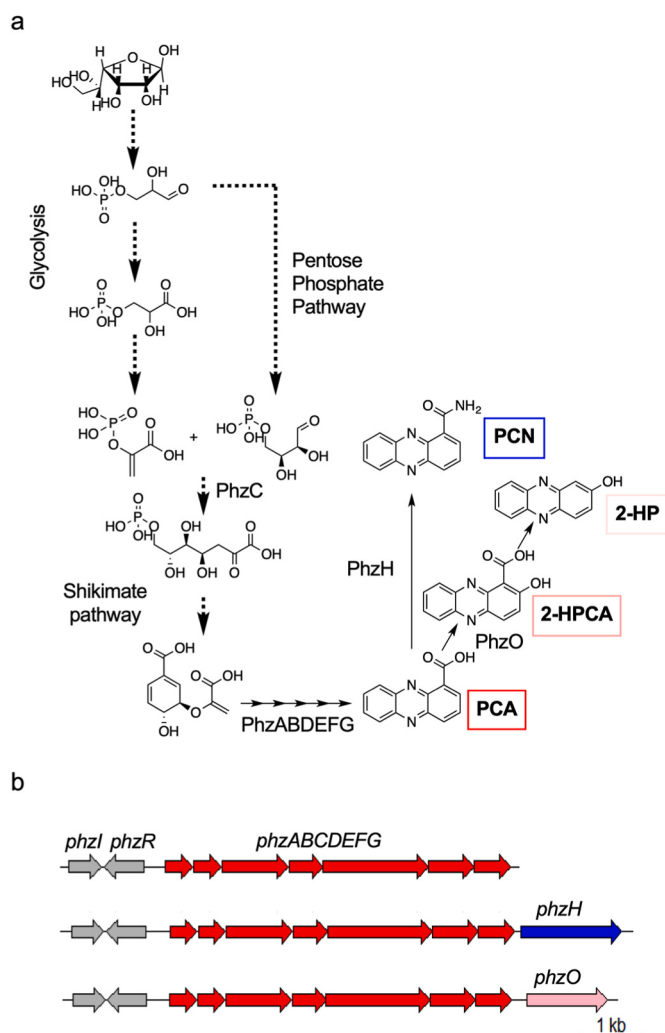


Fig. 1. Phenazine biosynthesis pathway and operon in *P. chlororaphis*. a) Phenazine biosynthesis pathway. *P. chlororaphis* naturally produces 4 phenazines: 2-hydroxyphenazine (2-HP), 2-hydroxyphenazine-1-carboxylic acid (2-HPCA), phenazine-1-carboxylic acid (PCA) and phenazine-1-carboxamide (PCN). Chorismate from the shikimate pathway is converted into PCA. PCA can be converted into PCN or 2-HPCA by PhzH or PhzO, respectively. 2-HP is a byproduct of the spontaneous decarboxylation of 2-HPCA. b) In fluorescent pseudomonads, phenazines are produced by a highly conserved core phenazine biosynthesis operon. *phzI*/*phzR* encodes a two-component quorum-sensing system which regulates the expression of the *phz* operon *phzABCDEFG*, which is responsible for the production of PCA. Some strains of *P. chlororaphis* contain *phzO*, the protein product of which converts PCA into 2-HPCA (which spontaneously decomposes into 2-HP), while others contain *phzH*, whose gene product converts PCA into PCN. *phzH* or *phzO* occur immediately downstream of *phzG* in the biosynthesis operon.

production of PCA and its derivatives. This strain originated from the Netherlands and was deposited into NCCB in 1988 (https://wi.knaw.nl/page/NCCB_strains_display/24262). This phenazine production data for both media conditions was used as input for the mGWAS analysis.

In addition to phenazine production, we also characterized growth temperature and biofilm formation for all strains in both media conditions and used these phenotypic datasets to assess the potential of each isolate as a biotechnology host (Thorwall et al., 2020). To identify strains which could grow at common bioprocessing temperatures, we characterized growth for all strains at 30 °C and 37 °C. These temperatures are relevant for *P. chlororaphis* since it is typically cultured at 28–30 °C, and some strains have been reported to grow at a maximum of

37 °C (Conway et al., 1956; Haynes and Rhodes 1962). While some other fluorescent pseudomonads like *P. aeruginosa* can grow well and produce phenazines at 37 °C, there are no reports of phenazine production in *P. chlororaphis* at this temperature. In order to measure growth at 30 °C and 37 °C, strains were grown on solid media (KMB and KMB + Fe) at the two temperatures and the opacity of colonies (i.e., the brightness of every colony pixel relative to that of the background, summed over the entire colony size) was measured after 48 h of growth (see Materials and Methods for experimental details). At 37 °C, the average opacity of strains was relatively lower compared to that at 30 °C, indicating worse growth at 37 °C (Fig. 2; see Fig. S1 for growth curves of all strains across 3 days in each condition). We also observed no colorful pigments on colonies at 37 °C, indicating little to no phenazine production, which agrees with the literature observations for *P. chlororaphis* (Conway et al., 1956). The strains which can grow at 37 °C could be useful to pursue as hosts for other products but likely not for phenazines. Therefore 30 °C was selected as the fermentation temperature for this study, as all isolates could grow at the lower temperature on both media.

Biofilm formation is a phenotype which affects a host's ability to produce a desired product through altered cellular metabolism and growth kinetics and can be engineered within pseudomonads (Benedetti et al. 2016). Given this, we characterized biofilm formation to determine whether future metabolic engineering efforts would be necessary to alter biofilm production in the desired host strain. For this study, we chose to minimize biofilm formation because high biofilm formation may be difficult to effectively culture in a planktonic system and/or clean from industrial bioreactors. In KMB, only 2 strains produced noticeable biofilm. Biofilm formation did increase in KMB + Fe, but only 5 strains had enough biofilm for it to be visibly noticeable when handling the liquid culture. Strain DSM 21509 and PCN were selected as the production strain and desired phenazine product, respectively, because this strain produced the highest overall phenazine titers, which were 99.2% PCN in KMB + Fe. Additionally, DSM 21509 had low biofilm formation in both tested media conditions making it favorable to work with.

2.3. Genome sequencing, assembly, and annotation

We sequenced all strains with both Illumina and Oxford Nanopore technologies as each technology generates reads which vary in length and accuracy, therefore affecting the quality of the resulting assemblies. Using each read set separately or together (in a hybrid approach), we assembled genomes with different assembly algorithms (i.e., SPAdes, Unicycler, Flye) to determine which algorithm and combinations of parameters yield the best assemblies. The summary statistics (i.e., number of contigs, L_{50} , N_{50} , assembly length, GC content, number of CDS and BUSCO score) were compared to assess genome contiguity and accuracy and thus select the optimal genome assemblies for the mGWAS analysis (Fig. S2).

Contiguity statistics (number of contigs, N_{50} , L_{50}) describe the degree of fragmentation of an assembly. The number of contigs, or assembly fragments, should ideally approach one to accurately represent bacterial genomes with a singular circular chromosome, as is expected for *P. chlororaphis*. Genome completeness was assessed by implementing the BUSCO algorithm to calculate the percentage of expected complete and single copy orthologs which are present in each strain. The hybrid assemblies created with Unicycler were selected as the final assemblies due to their high contiguity and completeness metrics ($L_{50} = 1$ and BUSCO score >98% for all strains). Summary statistics for each of the final genomes, including total length, number of annotated CDS, BUSCO scores, and assembly metrics are presented in Table 1. Notably, the vast majority of assemblies resulted in a single contig (22 *P. chlororaphis* and 1 *P. synxantha*), nine assembled into three or less contigs, one produced five contigs, and only one had ten contigs. Combined with the high BUSCO scores, the low number of contigs is indicative of high quality, complete genomes across our strain collection.

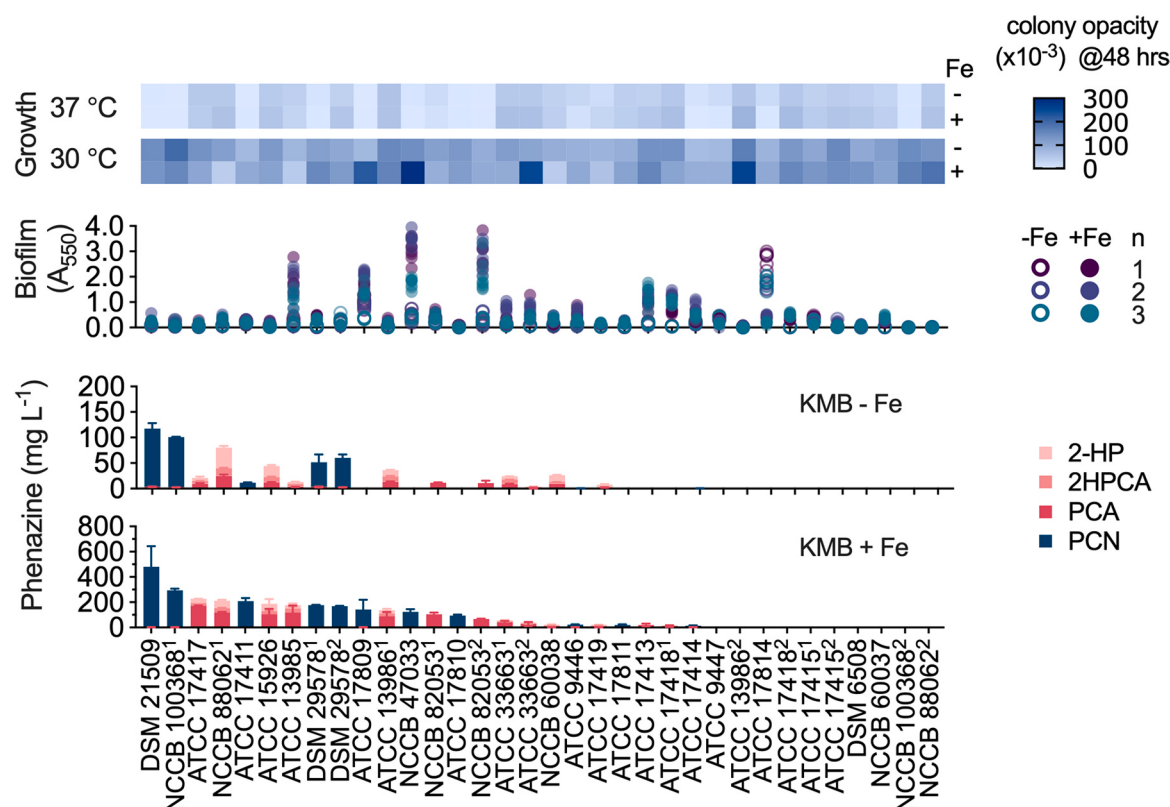


Fig. 2. Phenazine production, biofilm formation, and growth temperature phenotyping for all isolates used in this study. All phenotyping data was collected after 48 h of culture in either King's Media B (-Fe) or King's Media B + 100 μ M Fe^{3+} (+Fe). 2-hydroxyphenazine (2-HP), 2-hydroxyphenazine-1-carboxylic acid (2-HPCA), phenazine-1-carboxylic acid (PCA) and phenazine-1-carboxamide (PCN) were quantified using HPLC. The PCN-producers primarily produced PCN, with only very small amounts of the PCA precursor detected (< 5 mg/L). For phenazine production, bars indicate average of 3 replicates, and error bars represent one standard deviation. For growth temperature, the heat map shows a measure of colony growth on solid media (opacity ($\times 10^{-3}$); higher opacity indicates larger and more dense colonies). For biofilm formation, each data point (A_{550} , which is indicative of biofilm formation) represents a separate biological replicate, which is the average of 8 technical replicates.

2.4. Assembling the *P. chlororaphis* pangenome

While the 33 *P. chlororaphis* isolates are members of the same species, their gene content varies among isolates. Comparing the assembly summary statistics (Table 1) reveals a wide range of assembly size (6.6 Mbp to 7.9 Mbp) and number of CDS (5884–7222 annotated CDS), so we decided to assemble the pangenome to gain further insight into these genomic differences. We input the Prokka annotations for the final *P. chlororaphis* genomes into the algorithm PEPPAN (Zhou et al., 2020) to calculate the pangenome, the total gene content of the strains. The pangenome contains 11,527 total CDS and 4406 CDS common to all strains (Fig. 3a). This translates to 61–75% of the CDS in each genome being common to all strains, the core *P. chlororaphis* genome. The remaining CDS are members of the accessory genome (genes present in some strains and absent in others), which corresponds to the majority of this pangenome (7121 genes). These accessory genes are found in a relatively small number of strains, while all strains contain the 4406 core genes (Fig. 3b).

In addition to the genomic variation among the strains, we observed considerable phenotypic variation as shown in Fig. 2. Most notably, strains that produced PCN did not accumulate significant quantities of PCA or other phenazine derivatives. One of the key differences amongst these strains was the presence or absence of *phzH* and *phzO*; those that produced PCN contained *phzH*, while those that produced 2-HPCA contained *phzO*, as expected (Fig. 4). This split amongst the population is observable in a similarity tree based on accessory gene content, which clustered the isolates containing *phzH* separately from those that contain *phzO*. In the pre-genomic era, the phenotypic differences were

the basis of classification; members of these two groups would likely have been classified as different *Pseudomonas* species (e.g., green pigment-producers as *P. chlororaphis*, yellow-orange pigment-producers as *P. aureofaciens* or *P. aurantiaca* (Peix et al., 2007)). Classification in this way would incorrectly separate the groups, as our 16s and genomic sequencing shows that all 33 strains are *P. chlororaphis* with genetic variation driving the naturally produced phenazines. Collectively, the pangenome represents a large number of potential metabolic engineering targets that will be analyzed in our mGWAS.

2.5. Identifying phenazine biosynthesis hits by mGWAS

We used the hybrid genome assemblies and the phenazine production data to carry out an mGWAS for identification of genetic signatures associated with phenazine production. Phenazine biosynthesis was split into a series of phenotypes pertaining to production of PCA, PCN, and total phenazines in KMB and KMB + Fe for the mGWAS analysis (see Materials and Methods for the complete list of phenotypes). Table S3 shows the number of significant hits (i.e., unique DNA sequences or unitigs) obtained for each phenotype in the mGWAS. All hits and their reverse-complement sequences were aligned to genomes of all strains to find their genomic locations, resulting in a 'preliminary list' of 2493 significant hits across all phenotypes (Fig. 5a, Supplementary File 1). Each unitig (and its reverse-complement) in this list may be found in one or more strains and in one or more phenotypes, thus creating redundancies in the list as the unitigs were counted multiple times. These redundancies were eliminated by collapsing the hit list in 3 stages – phenotype-collapsing, strain-collapsing, and reverse-complement

Table 1

Summary statistics for final genome assemblies. For each isolate, the assembly length, number of contigs, N₅₀, L₅₀, GC% and BUSCO score are reported. These assemblies were used as input to mGWAS analysis. The number of CDS was tallied from the Prokka genome annotations, and the BUSCO score was calculated as the percentage of complete and single copy BUSCOs present in each genome using the BUSCO algorithm. All other statistics were generated from QUAST. These genome assemblies are available at NCBI with the listed accession numbers. All strains are *P. chlororaphis*, except ATCC 17413 (marked with *), which is *P. synxantha*.

Strain	Total length (bp)	Contigs	N ₅₀	L ₅₀	GC (%)	CDS	BUSCOs (%)	NCBI Accession Number
ATCC 13985	7 024 010	10	4 636 000	1	62.7	6251	99.1	JAQZQZ000000000
ATCC 13986 ¹	6 675 284	2	6 636 555	1	63.0	5926	99.5	JAQZQY000000000
ATCC 13986 ²	6 682 756	2	6 644 045	1	63.0	5949	99.5	JAQZQX000000000
ATCC 15926	6 763 921	1	6 763 921	1	62.9	5998	99.4	CP118156
ATCC 17411	7 212 419	1	7 212 419	1	62.5	6366	99.2	CP118155
ATCC 17414	6 807 169	1	6 807 169	1	63.0	6048	99.5	CP118154
ATCC 17415 ¹	6 664 157	1	6 664 157	1	63.0	5887	99.4	CP118147
ATCC 17415 ²	6 664 503	1	6 664 503	1	63.0	5884	99.2	CP118146
ATCC 17417	6 746 536	1	6 746 536	1	62.9	5954	99.1	CP118145
ATCC 17418 ¹	6 883 267	1	6 883 267	1	62.8	6075	99.5	CP118144
ATCC 17418 ²	6 881 643	1	6 881 643	1	62.8	6074	99.5	CP118143
ATCC 17419	6 608 598	5	4 662 896	1	62.7	5919	99.0	JAQZQW000000000
ATCC 17809	7 020 903	1	7 020 903	1	62.4	6223	99.0	CP118142
ATCC 17810	6 863 056	2	6 791 445	1	62.7	6074	99.4	JAQZQV000000000
ATCC 17811	7 189 114	1	7 189 114	1	62.4	6422	99.4	CP118153
ATCC 17814	6 807 913	1	6 807 913	1	63.0	6050	99.4	CP118141
ATCC 33663 ¹	7 109 352	1	7 109 352	1	62.9	6281	99.1	CP118152
ATCC 33663 ²	7 108 820	1	7 108 820	1	62.9	6284	99.2	CP118140
ATCC 9446	6 637 791	1	6 637 791	1	63.0	5909	99.2	CP118151
ATCC 9447	6 807 068	3	6 677 872	1	63.0	6048	99.4	JAQZQU000000000
DSM 21509	7 064 975	1	7 064 975	1	62.7	6246	99.1	CP118150
DSM 29578 ¹	7 216 947	1	7 216 947	1	62.5	6378	99.1	CP118139
DSM 29578 ²	7 216 571	1	7 216 571	1	62.5	6380	99.1	CP118138
DSM 6508	7 915 166	3	7 476 725	1	62.5	7222	99.4	JAQZQT000000000
NCCB 100368 ¹	6 870 522	2	6 455 838	1	62.8	6010	99.2	JAQZQS000000000
NCCB 100368 ²	6 870 415	2	6 455 628	1	62.8	6010	99.1	JAQZQR000000000
NCCB 47033	7 221 530	1	7 221 530	1	62.4	6378	99.2	CP118137
NCCB 60037	6 977 278	1	6 977 278	1	62.7	6209	99.1	CP118149
NCCB 60038	6 979 353	1	6 979 353	1	62.7	6214	99.1	CP118136
NCCB 82053 ¹	6 763 242	2	6 274 577	1	62.9	6002	99.4	JAQZQQ000000000
NCCB 82053 ²	6 762 156	1	6 762 156	1	62.9	6002	99.4	CP118135
NCCB 88062 ¹	7 025 460	2	6 660 177	1	62.8	6233	99.1	JAQZQP000000000
NCCB 88062 ²	6 923 225	1	6 923 225	1	62.8	6138	98.2	CP118148
*ATCC 17413	6 147 644	1	6 147 644	1	60.0	5459	99.9	CP118134

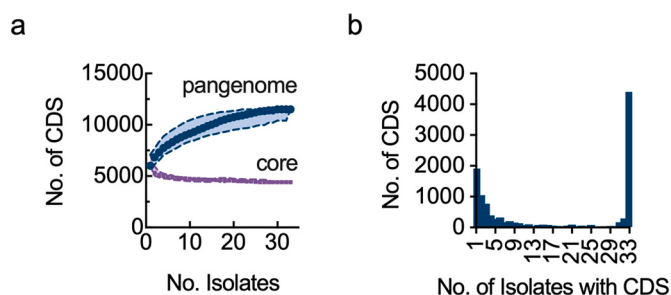


Fig. 3. Summary of the pangenome constructed from the final hybrid genome assemblies. (a) The pangenome rarefaction curve shows how the total pangenome size (blue) increases and the core genome size (purple) decreases as isolates are added to the pangenome. (b) Histogram showing how many CDS are only found in a specific number of strains. Core genes present in almost all strains or accessory genes only found in a few are present with the highest frequencies.

collapsing – to result in a ‘final list’ that only contains a unique entry for each significant hit influencing phenazine biosynthesis. Phenotype collapsing reduced the list to 1568; strain collapsing condensed the list further to 474. Finally, removing entries that were due to reverse complement redundancy resulted in a final list of 330 unique genomic hits for phenazine biosynthesis. The corrected p-value of the hits in the final list is shown in Fig. 5b. Fig. S3 illustrates the collapsing pipeline, and Supplementary File 2 contains all entries in each of the collapsed lists. A vast majority of the hits (284 out of 330) have a positive effect on phenazine production while the remaining ones had a negative effect (Fig. S4).

To visualize the genomic location of hits in the final list, we created a circos plot and mapped as many hits as possible to a PCN- or PCA-producing strain (Fig. 5d). Since each strain contains a different set of accessory genes and many of the hits map to the accessory genome, we were unable to map all hits to a single strain. We selected NCCB 100368¹ as the basis for displaying the PCN-related hits because it contains a majority of the PCN hits (122 out of 127). Similarly, NCCB 88062¹ was selected as the basis to display PCA related hits as it contained 170 out of 203 PCA-related hits, more than any other strain. In total, the two strains combined contain 292 hits from the final list (out of 330). Mapping the hits revealed that there is little association between the PCA hits and PCN-producing strains. Furthermore, the two hits related to total phenazine production mapped to PCN-producing strains only. In comparison, many of the hits related to PCN biosynthesis were found in both PCA- and PCN-producing strains.

The mGWAS analysis links unitigs to phenazine production, the next step in our analysis is to identify which genes are associated with each unitig. This mapping is straightforward when the unitig partially or completely overlaps to a CDS, but the connection to a gene is less clear when the unitig is contained within an intergenic region. In these cases, we identified genes upstream and downstream of the intergenic region as potential genes of interest. In total, the 330 unitig hits map to 158 genes in the pangenome (many of the genes were associated with more than one unitig; see Supplementary File 3). The 158 genes include 80 functionally annotated proteins or homologs of proteins with known function and 78 hypothetical proteins. While the function of the hypothetical proteins are unknown, 33 of them belonged to the core genome while the remainder belonged to the accessory genome. Forty-one out of the 80 proteins of known function belonged to the *P. chlororaphis* core genome.

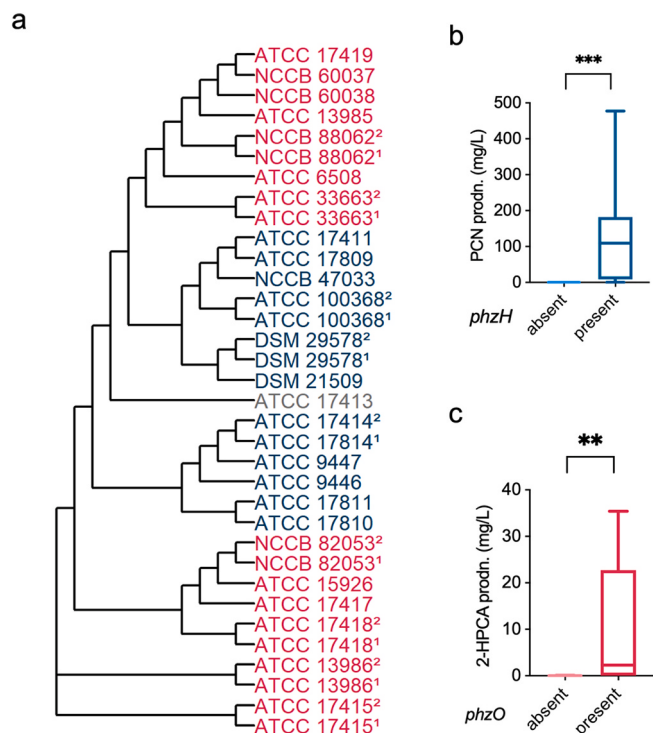


Fig. 4. Categorizing isolates based on presence and absence of genes *phzH* and *phzO*. (a) Tree based on accessory gene content among the *P. chlororaphis* pangenome. Strains containing *phzO* (red) and *phzH* (blue) form groups with similar gene content. (b), (c) Box-and-whisker plots showing significantly higher (**p < 0.001) PCN production in strains containing *phzH* and significantly higher (**p < 0.01) 2-HPCA production in strains containing *phzO* in KMB + Fe based on one-tailed unpaired *t*-test. Phenazine production data in (b) and (c) is taken from the phenotype data shown in Fig. 2.

In the phenotype-collapsing stage we associated each hit to a 'phenotype score', which represents the number of phenotypes in which a hit was found (Fig. S3); the higher the score, the greater the number of phenotypes influenced by the hit. The phenotype score was used as a metric to identify hits and associated genes most likely to improve phenazine biosynthesis. Fig. 5c shows the distribution of phenotype scores for the final list. We deemed hits with a phenotype score of 3 or higher as most likely to affect phenazine biosynthesis, and therefore ones that we sought to target for further analysis. The top phenazine-producing strain DSM 21509 primarily produces PCN, and the hits influencing PCN production have lower *q*-values than those influencing PCA production (Fig. 5b). We therefore narrowed our mGWAS validation and metabolic engineering studies to PCN producing hits only. Table 2 shows the attributes of these hits, two of which have a phenotype score of 4, while the other two have a score of 3. All of these hits were found to be single nucleotide polymorphisms (SNPs) in the coding or intergenic (non-coding) regions. Genes containing or adjacent to SNPs for PCN production include YbhH (a putative isomerase), RhtA (threonine/homoserine exporter), UctC (acetyl-CoA:oxalate CoA-transferase), ProY_1 (proline-specific permease), HutH2 (histidine ammonia-lyase), and two hypothetical proteins (annotated as PS_04251 and PS_04252 in DSM 21509), all of which belong to the *P. chlororaphis* core genome. DSM 21509, the highest PCN-producing strain, contains 3 of the 4 SNPs for PCN production.

2.6. Validating mGWAS hits for PCN production

We overexpressed the top gene hits for PCN production in DSM 21509 to verify their phenotypic effects. Three of the genes yielded significant changes in PCN production when overexpressed in DSM 21509. Overexpression of PS_04252 in KMB produced 19.1 ± 3.5 mg/L of PCN compared to 58.7 ± 12.3 mg/L produced by the empty vector control, which is a 67.5% decrease (Fig. 6a). This target whose overexpression decreases PCN production could provide a target for knockout to increase PCN production. Overexpressing ProY_1 and

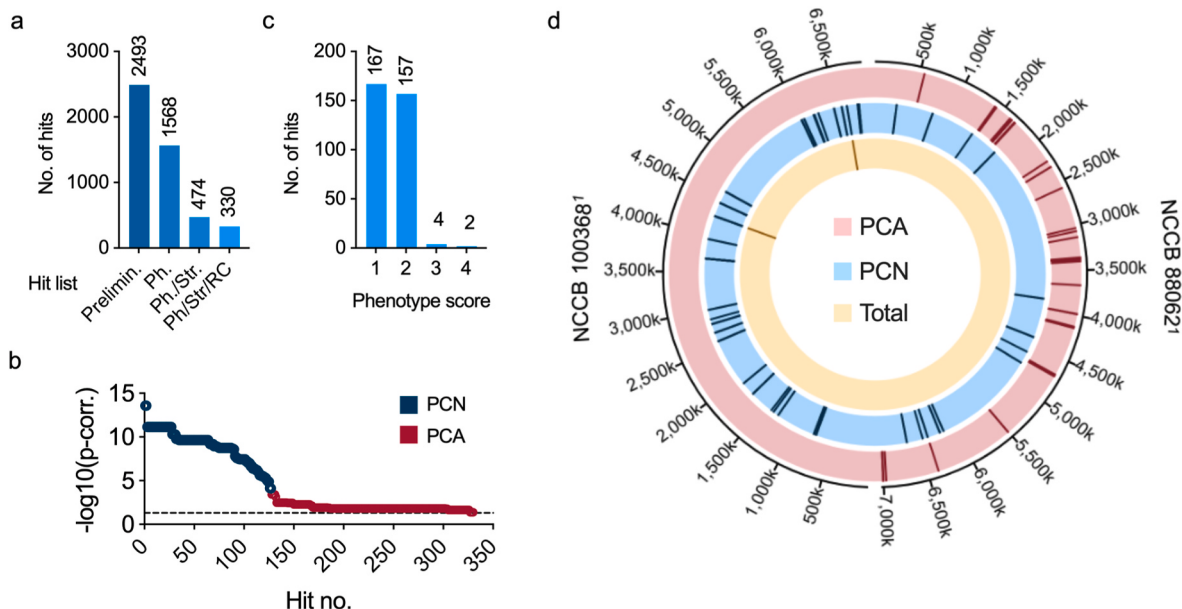


Fig. 5. Results of the mGWAS analysis for phenazine production. (a) Number of significant mGWAS hits in the preliminary (uncollapsed) list and lists obtained after each collapsing stage - phenotype-collapsed list (Ph), phenotype + strain-collapsed list (Ph./Str.), and phenotype + strain + reverse complement-collapsed list (Ph./Str./RC; also called the 'final list'). Numbers above each bar indicate the exact number of hits in the list corresponding to that bar. (b) Corrected *p*-values of the 330 hits in the final list. Hits were numbered in decreasing order of $-\log_{10}(\text{p-corrected})$ value, and were grouped into those influencing PCA production and PCN production. (c) Phenotype score distribution of hits in the final list. Numbers above each bar indicate the total number of hits having phenotype score corresponding to that bar. (d) Circos plot showing genomic locations of hits in the final list grouped into 3 categories based on the phenotype(s) in which they were significant: PCA production, PCN production, and total phenazine production, with respect to 2 strains - NCCB 100368¹ and NCCB 88062¹.

Table 2

Most influential hits for PCN production, identified in the final list of significant mGWAS hits. Hits are numbered 1 through 4, and attributes such as phenotype score, associated phenotypes, strains, and genes, as well as variant type, p-value, effect on phenotype (positive/negative), & genomic region have been provided for each hit.

No.	Phenotype score	Major phenotypes	Associated gene(s)	Corr. p-value (effect)	Genomic region	Type of variant	Strains
1	4	PCN production	hypothetical protein (PS_04251)	2.6×10^{-14} (+)	intergenic	SNP	DSM 21509, DSM 29578 ¹ , DSM 29578 ² , NCCB 100368 ¹
2	4	Total phz. production	hypothetical protein (PS_04252)	2.6×10^{-14} (+)	CDS	SNP	DSM 21509, DSM 29578 ¹ , DSM 29578 ² , NCCB 100368 ¹
3	3	PCN production	Putative isomerase (YbhH)	6.7×10^{-10} (–)	intergenic	SNP	DSM 29578 ¹ , DSM 29578 ² , NCCB 100368 ²
4	3	PCN production	Threonine/homoserine exporter (RhtA) Acetyl-CoA:oxalate CoA-transferase (UctC) Proline-specific permease (ProY_1) Histidine ammonia-lyase (HutH_2)	2.2×10^{-9} (+)	intergenic	SNP	DSM 21509, DSM 29578 ¹ , DSM 29578 ² , NCCB 100368 ¹ , NCCB 47033, ATCC 17411, ATCC 17809, ATCC 17414, ATCC 9446, ATCC 17814, ATCC 9447, ATCC 17811, ATCC 17810

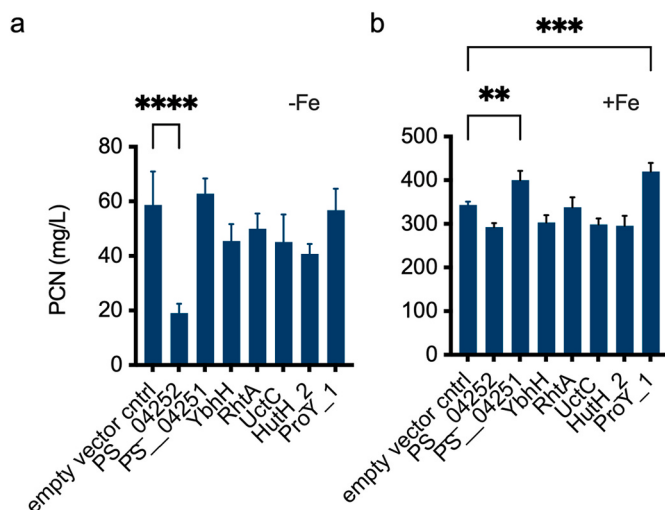


Fig. 6. Overexpressing top hits in the highest PCN-producing isolate, DSM 21509. Genes associated with top phenotype-scoring units for PCN production were overexpressed in the top PCN-producing strain DSM 21509. PCN was quantified after 48 h of culture in (a) King's Media B and (b) King's Media B + Fe. Bars represent the average of 3 replicates and error bars represent one standard deviation. Asterisks denote p-values <0.01 (**), <0.001 (***) and <0.0001 (****) when performing an ordinary one-way ANOVA followed by Dunnett's test with respect to the empty vector control.

PS_04251 increased PCN production in KMB + Fe to 420.2 ± 19.7 mg/L and 400.1 ± 21.5 mg/L, respectively, compared to the 343.6 ± 7.3 mg/L PCN produced by the empty vector control (Fig. 6b). The hits validated here could be combined with other known beneficial genetic manipulations and/or applied to previously engineered and process-optimized strains (Li et al., 2020: 11.45 g/L PCN) to further improve phenazine production.

PS_04251 and PS_04252 were both associated with unitig 1. PS_04252 is annotated as a hypothetical protein and has 57% sequence identity and 97% coverage to helix-turn-helix (HTH) MarR-family transcriptional regulator PA1607 from *P. aeruginosa* (NCBI Reference Sequence NP_250298.1) (Kaur and Subramanian 2015). The MarR, or multiple antibiotic resistance repressor, family often regulates expression of multidrug efflux pumps and some, including PA1607, may derepress in response to oxidative stress (Housseini B Issa et al., 2018; Kaur and Subramanian 2015). Therefore overexpression of the putative

regulator PS_04252 could lead to increased repression of its target which could explain the observed decrease, rather than increase, in PCN production. PS_04251 is a hypothetical protein with unknown function which has 86% identity and 100% coverage to putative M14-type zinc cytosolic carboxypeptidase PSF113_3889 (NCBI Reference Sequence WP_041476041.1) from *Pseudomonas ogarae* whose function is unknown (Rimsa et al., 2014). While the function of this and similar proteins are unknown, we found its overexpression to significantly improve phenazine production in *P. chlororaphis*. Because both CDS surrounding unitig 1 were successful, this genomic region could be of further interest to investigate for phenazine production.

The other hit which significantly improved PCN production, ProY_1, was associated with unitig 4 which occurred within the hut operon, which is responsible for histidine catabolism. Due to its position in the highly conserved operon and its sequence similarity, the hit annotated as ProY_1 is likely the histidine permease HutT which imports histidine and is required for its utilization (X.-X. Zhang et al., 2012). One study suggests that histidine catabolism in *P. fluorescens* is connected to oxidative stress response, as it could increase intracellular pools of the antioxidant α -ketoglutarate (Lemire et al., 2010). Possibly, overexpressing HutT could contribute to improved oxidative stress tolerance through a similar mechanism in *P. chlororaphis* by increasing intracellular histidine levels and therefore α -ketoglutarate levels. The other CDS adjacent to unitig 4 is HutH_2 (HutH2), the histidine ammonia lyase which catalyzes the first step of histidine catabolism, the conversion of L-histidine into urocanate (X.-X. Zhang and Rainey 2007). While this hit is associated with the same operon, its overexpression did not affect PCN titers. Because unitig 4 was within the non-coding region between 2 CDS, the mGWAS results in actuality may have been connected to one CDS and not the other. By overexpressing both hits, we were able to identify the one which was relevant for phenazine production.

While overexpressing the genes associated with unitigs 2 and 3 did not produce an observable change in phenotype, they also appear to be related to oxidative stress as well as amino acid export and catabolism. Unitig 2 occurred within the gene annotated as putative isomerase YbhH. This CDS has 34% identity and 92% coverage to *E. coli* YbhH, which does not have a known function but its expression has been upregulated in response to the σ^E stress signaling pathway (Bury-Moné et al., 2009). Unitig 3 occurred within a non-coding region which was flanked by 16s rRNA and either UctC or RhtA. The hit annotated as UctC has 87% sequence identity and 100% coverage to *P. putida* KT2440 glutamate-CoA transferase GcoT (NCBI Reference Sequence NP_742328.1) which is part of L-lysine catabolism (M. Zhang et al., 2019). The other associated hit has 57% identity and 93% coverage to

E. coli RhtA (NCBI Reference Sequence WP_001 295 297), which exports threonine and homoserine in addition to other amino acids. While the mGWAS analysis identified these unitigs as significantly related to phenazine production, the overexpression studies showed that they did not alter PCN titers in DSM 21509.

Traditional approaches to metabolic engineering involve manipulation of native or engineered pathways in microbial hosts to direct metabolic flux towards synthesis of a target biochemical (Cho et al., 2022). The development of high-throughput technologies, such as next-generation sequencing (NGS), that generate massive amounts of data has made it possible to look for genetic targets spanning the entire genome, thereby facilitating strain engineering for various phenotypes of industrial importance. So far, this population genomics approach for metabolic engineering has been mainly applied to model organisms like *Saccharomyces cerevisiae* for design of strains manifesting higher biochemical production and higher tolerance to growth inhibitors present in the feedstock (Hubmann et al., 2013; Maurer et al., 2017; Meijnen et al., 2016). These studies exploit the natural diversity of strains to identify non-intuitive genetic variants underpinning the trait of interest. In this study, we sought to extend this approach to the production of phenazine compounds in the non-model bacterium *P. chlororaphis*. We used a collection of 34 *Pseudomonas* strains that were isolated from various environmental locations to ensure sufficient genomic diversity. We exploited this diversity within the strain collection to identify 330 variants within the pangenome that influence phenazine production. These pangenome-wide variants were associated to 158 genes, which serve as potential metabolic engineering targets for increasing phenazine biosynthesis in *P. chlororaphis*. To validate our metabolic engineering approach, we selected 7 gene hits and overexpressed them in DSM 21509, the strain with the highest PCN titers. We used the phenotype score to prioritize these hits over others because this method was unbiased in that it is blinded to gene function. Overall this data-driven approach was successful because we identified two candidate genes that improved phenazine production and one that reduced it.

An alternative approach is to prioritize hits based on a rational design strategy, that is, target genes with known functions related to the phenazine biosynthesis or associated pathways. For example, our hit list also includes GacA, a global transcriptional regulator known to impact phenazine production (Li et al., 2020; D. Wang et al., 2013; Supplementary File 3). Pursuing this target or others associated with the Gac regulatory cascade could be promising for phenazine production, as the mGWAS and literature results are in agreement. Moreover, since we verified successful hits that may be transcriptional regulators or are involved in oxidative stress response, hits from the list with similar functions could be prioritized for future studies. The other hits on the list which are annotated as HTH-transcriptional regulators (e.g., ArgP, BenM, CynR, MtrA, and RhaS) or the hits known to be connected to oxidative stress response from the literature (e.g., RscC (Bury-Moné et al., 2009), glutathione synthase (Nikel et al., 2021; Wongsaroj et al., 2018) and glucose-6-phosphate 1-dehydrogenase (zwf) (Kim et al., 2008)) could be pursued as targets in future studies. Importantly, our population genomics approach to metabolic engineering and the hits it generates could be used in tandem with other successful genetic targets from the literature and process optimization strategies to achieve additional improvements in product titers.

Many of the strains in our collection were isolated from diverse environmental locations to ensure natural phenotypic and genomic variation. In our analysis, we observed a broad range of phenotypes and genotypes that were distributed so that both positive and negative groups were well-represented. For example, similar numbers of strains contained *phzH* (14 strains) vs. *phzO* (19 strains), and similar numbers of strains produced no phenazines (10 strains), less than 100 mg/L phenazines (11 strains), and greater than 100 mg/L phenazines (13 strains). This phenotypic diversity, along with the genetic diversity found in the pangenome, allowed us to use a relatively small number of strains to perform the mGWAS analysis and obtain significant hits. The statistical

power of hit identification can be potentially improved by using a bigger and more diverse collection of isolates as input to the mGWAS analysis, entailing a larger representation of phenotypic and genomic diversity. Further, GO and pathway enrichment tests of genes associated with 330 hits resulted in no enriched terms (see Supplementary File 4, Supplementary File 5). Pursuing hits based off their functions will become more promising as more microbes are annotated for GO-terms and metabolic pathways.

Taken together, this work presents a new approach that enables genome-scale metabolic engineering of pseudomonads. This approach is data-driven, using the power of low cost sequencing and high throughput phenotyping to generate large data sets that correlate desired traits to genomic variants within a microbial population, thereby generating new metabolic engineering targets. While we demonstrate this approach in *Pseudomonas*, it can be extended to other microbial species, especially non-conventional microbes exhibiting industrially relevant phenotypes as new strains are discovered and whole genome sequences become available. Upon gene target elucidation, the microbial strains can be engineered to improve biochemical production or tolerance to various environmental stresses inhibiting cell growth, among other phenotypes. Identification of genetic targets for engineering more complex phenotypes could be accomplished by using a collection of isolates belonging to different but related species, resulting in greater genetic diversity and hence, a more complex pangenome. While rational design strategies for many of these phenotypes may have been previously developed, novel hits identified using a population genomics approach could be used in conjunction with those to further enhance the phenotype of interest and consequently scale-up industrial bioprocesses.

3. Conclusions

Advancements in whole genome DNA sequencing and genome-editing techniques, as well as increased availability of bioinformatics tools for analysis of genome-wide data have allowed us to identify metabolic engineering targets spanning the entire pangenome. The accessory genome and core genome are promising sources of metabolic engineering targets for the bacterial production of secondary metabolites such as phenazines. The present study taps into both of these pangenome components to help identify strain engineering targets for biosynthesis of the phenazine PCN in *Pseudomonas chlororaphis*. This pangenome-wide approach, in combination with rational design approaches, could potentially lead to substantial improvement in the phenotype of interest, while also assisting with selection of the appropriate host strain for metabolic engineering.

4. Materials and methods

4.1. Strain selection and culturing

All strains designated as *Pseudomonas chlororaphis* that were available as of April and October 2019 were ordered from the American Type Culture Collection (ATCC; Manassas, VA); all strains designated as *Pseudomonas chlororaphis* that were nonredundant and available as of March 2020 were ordered from the German Collection of Microorganisms and Cell Cultures (DSMZ GmbH; Braunschweig, Germany) and the Westerdijk Fungal Biodiversity Institute's Netherlands Culture Collection of Bacteria (NCCB; Utrecht, Netherlands). Strains which appeared to have more than one colony morphology present were separated into distinct isolates (denoted by ¹ and ², which were arbitrarily assigned) that were sequenced and cultured separately. All isolates were sequenced with 16s rRNA sequencing (GENEWIZ®; South Plainfield, NJ), and the 33 confirmed *P. chlororaphis* isolates were used in this study (Table S1). One PCA-producing *P. synxantha* isolate was also used in this study as a phylogenetic outgroup for a total of 34 isolates.

Strains were initially revived according to the guidance of each

culture collection then subsequently cultured at 30 °C in King's Media B (KMB), the standard media for fluorescent pseudomonads culture, according to the methods of King et al. (King et al. 1954). To improve phenazine production, KMB + Fe Media was made by supplementing KMB with 100 µM ferric sodium ethylenediaminetetraacetate (FeNaEDTA) based off the findings of van Rij et al. (van Rij et al. 2004). For phenazine production experiments, both KMB and KMB + Fe contained 1.5 g/L MgSO₄. Cultures were supplemented with 50 µg/mL kanamycin sulfate when an antibiotic resistance marker was used. Luria Bertani (LB) broth and TOP10 chemically competent *E. coli* cells were used for cloning.

Liquid culturing was performed using sterile 2 mL 96-deep well plates within an INFORS HT Multitron Pro plate shaker incubator at 1000 rpm and ~88% humidity. Overnight cultures were started by inoculating 500 µL media of interest with the respective colony or glycerol stock. After the overnight culture was incubated with shaking at 30 °C for 22–24 h, the plate was spun down in Beckman Coulter Allegra 25 R centrifuge for 10 min at 5,000 g. To reduce phenazine transfer and to ensure biofilm-forming strains were well-mixed, old media was removed, and cultures were resuspended in fresh media. To start experimental cultures, 500 µL of desired media were inoculated with 10 µL of resuspended overnight culture. For cultures requiring induction with isopropyl β-D-1-thiogalactopyranoside (IPTG), sterile-filtered IPTG was added to cultures to a final concentration of 1 mM about 4 h after inoculation.

4.2. Phenazine quantification

48 h after inoculation, phenazine compounds were extracted from each liquid culture using ethyl acetate liquid-liquid extraction. Whole cultures were acidified with 10 µL of 3 M HCl, then 1.2 mL ethyl acetate was added to each culture. Each mixture was transferred to a micro-centrifuge tube, vortexed at maximum speed for 1 min, and spun down to separate liquid phases. The ethyl acetate phase was evaporated, resuspended in methanol, and filtered for quantification via HPLC.

All phenazines were quantified with a photodiode-array detector on a Shimadzu Nexera-i LC-2040C 3D liquid chromatograph with an Agilent Poroshell 120 EC-C18 2.7 µm 3.0 × 75 mm column and 3.0 mm × 5.0 mm guard column at 40 °C. To resolve the similar phenazine derivatives, the following method with gradients of methanol and ammonium acetate buffer (pH 5.0) was used: 2 µL sample injection, 5 min of 20% methanol, 2 min gradient from 20% to 30% methanol, and 13 min gradient from 30 to 40% methanol with subsequent steps to wash and re-equilibrate the column, with all steps at a 1 mL/min flow rate. PCA and PCN peaks were identified by comparing retention times to those of purchased PCA and PCN (ChemScene; Monmouth Junction, NJ). Because pure 2-HP and 2-HPCA were not commercially available, the identities of these HPLC peaks were confirmed with LC-MS following the same protocol. Phenazines were quantified by converting peak areas at a wavelength of 254 nm and bandwidth of 4 nm to concentrations using extinction constants calculated from the purchased PCA and PCN.

4.3. Biofilm formation and growth temperature phenotyping

Biofilm formation phenotyping was characterized using a crystal violet staining assay following the protocol of O'Toole (O'Toole 2011). To adapt the protocol for *Pseudomonas chlororaphis*, a 2% inoculum of overnight culture in the respective media was used to start stationary cultures which were incubated without agitation at 30 °C for 48 h.

For growth phenotyping, *Pseudomonas* cultures were grown in KMB for 48 h, diluted to an OD₆₀₀ of 1, then 1 µL was transferred onto respective solid media and incubated at two different temperatures: 30 °C and 37 °C. Four technical replicates were performed for each sample. The plates were imaged every 24 h for 3 days with an Epson V850 scanner and the images were processed using Iris v0.9.7 (Kritikos et al., 2017; mode: Colony Growth). Any colonies that were missed were

reprocessed using the ColonyPicker software provided with Iris. Similar to previous studies (Banzhaf et al., 2020; Ropars et al., 2020), the opacity of colonies was used as an indicator for colony growth. Average opacity for each strain was calculated as the mean of opacity values across all replicates. The value of average opacity on day 2 (i.e., after 48 h of growth) was used to gauge the ability of strains to grow at the respective temperature levels. Opacity values of all strains in each condition for days 1–3 have been provided in **Supplementary File 6**.

4.4. Genome assembly

Genomic DNA was isolated using the Quick-DNA™ Fungal/Bacterial Miniprep Kit (Zymo Research; Irvine, CA) and sent to the Microbial Genome Sequencing Center (Pittsburgh, PA) for whole genome sequencing. All isolates were sequenced on the NextSeq 2000 (Illumina; San Diego, CA) with paired-end 150 base pair reads and with Oxford Nanopore technologies. Illumina read quality was assessed using FASTQC v0.11.9 (Andrews and Others 2010) and Nanopore read statistics were assessed with NanoStats v1.28.2 (De Coster et al., 2018) on the Galaxy platform before and after read filtering and trimming (the parameters for all bioinformatics tools are available in Table S2). Summary statistics (i.e., total assembly length, number of contigs, N₅₀, L₅₀, % GC) were calculated using QUAST v5.0.2 (Mikheenko et al., 2018). Genome completeness was assessed by running BUSCO v5.2.2 in genome mode using the pseudomonadales_odb10 (prokaryota, 2020-03-06) database (Manni et al., 2021).

Flye genome assemblies were created with Flye v2.8.3 and raw Nanopore reads as input (Kolmogorov et al., 2019). All other genome assemblies used reads which were filtered and trimmed based on read quality. Raw Illumina reads were trimmed to remove adapters and low-quality ends using Trimmomatic v0.38 (Bolger et al. 2014). Raw Nanopore reads were adaptor-trimmed using Porechop v0.2.4 (R. R. Wick 2018) then filtered with filtlong v0.2.1 (R. Wick and Menzel 2019). SPAdes genome assemblies were created using SPAdes v3.12.0 (Prjibelski et al., 2020) on the Galaxy platform (Afgan et al., 2018); Unicycler assemblies were created using Unicycler v0.4.8 using both “Normal” and “Bold” bridging modes and excluding contigs shorter than 1000 bp from the assemblies (Ryan R. Wick et al., 2017). The short-reads assemblies were created using only the paired end Trimmomatic output. The long-reads Nanopore assemblies were created using the trimmed and filtered Nanopore reads. The hybrid assemblies were assembled using both sets of aforementioned reads.

4.5. Genome annotation and pangenome assembly

Genome assemblies were annotated with Prokka v1.14.6 (Seemann 2014) on the Galaxy platform (Afgan et al., 2018), using a minimum contig size of 1000 and ‘Pseudomonas’ as the genus name. Prokka outputs genome annotations in GFF3 format. These GFF3 files were used along with the draft genome assemblies to generate annotated genome FASTA files by bedtools GetFastaBed v2.30.0 (Quinlan and Hall 2010). The pangenome was constructed using PEPPAN v1.0.5 and the gff files generated by Prokka as input (Zhou et al. 2020). A rarefaction curve, gene presence absence matrix, and accessory genome tree were created from the PEPPAN output using the included PEPPAN_parser algorithm. Statistics about the core genome were calculated from the gene presence absence matrix using R v4.2.1 (RStudio, 2022.07.1). The resulting .nwk tree file was visualized using R v4.2.1 and treeio package v1.20.2 (L.-G. Wang et al., 2020). All remaining figures were created using GraphPad Prism v9.4.1 (GraphPad Software; San Diego, CA).

4.6. Genome-wide association study

De novo assembled genomes of the 34 *Pseudomonas* strains were provided as input to DBGWAS v0.5.4 (Jaillard et al., 2018) along with the corresponding phenotype values for phenazine production.

DBGWAS was implemented for 7 different phenotypes: (i) PCA production in KMB; (ii) PCA production in KMB + Fe; (iii) Effect of Fe on PCA production; (iv) PCN production in KMB; (v) PCN production in KMB + Fe; (vi) Effect of Fe on PCN production; (vii) Total phenazine production in KMB.

For phenotypes (i), (ii), (iv) and (v), concentrations of PCA and PCN (mg/L) obtained from experiments were used directly. Values of phenotypes (iii) and (vi) were estimated by subtracting the concentration of the phenazine compound in KMB from that in KMB + Fe. If the difference was negative, it was replaced by 0. Total phenazine production in KMB was obtained by simply adding the concentrations of all phenazine compounds (*i.e.*, PCA, 2-HPCA, PCN and 2-HP) in KMB for each strain. The genome sequences of strains were also provided as BLAST database to DBGWAS for genome mapping of significant unitigs. Significant unitigs were identified based on a corrected p-value cutoff, and a minor allele frequency greater than 1% (default). The number of significant unitigs obtained for each phenotype are listed in Table S3.

4.7. Downstream processing of mGWAS hits

Even though DBGWAS maps significant unitigs to genomes by BLAST, we chose to independently perform unitig alignment to genomes by exact matching to avoid any tolerance to mismatches during alignment. Custom Python3 scripts were used for this purpose with the 34 genome sequences as the mapping database. To ensure that each unitig finds a match, both the significant unitig and its reverse-complement were used. Further, genome annotations were used to determine the genomic regions of the mapped unitigs (*i.e.*, whether the unitig falls within a gene or an intergenic region).

The lists of mGWAS hits from the 7 phenotypes were concatenated into a single list called the ‘preliminary list’. In this list, each occurrence of a significant unitig constituted a single entry, creating separate entries for each phenotype, strain, as well as the reverse complement sequence of that unitig. Custom Python3 scripts were used to remove redundancies and collapse the preliminary list so that each significant unitig has a single entry in the final list (Fig. S3). For each strain, entries for identical unitigs that were significant for multiple phenotypes were collapsed together in the ‘phenotype-collapsed’ list. Each entry was assigned a phenotype score, which represents the number of phenotypes (out of 7) where each unitig was significant. If a unitig had a phenotype score greater than 1, its corrected p-value was taken to be the minimum of corrected p-values for all phenotypes it is found in. Similarly, the effect of that unitig was taken to be the one with the highest magnitude across all phenotypes. The ‘phenotype + strain-collapsed list’ combined entries where the same unitig mapped to the same genomic region in multiple strains. Redundancies where the reverse complement of a significant unitig shows up as a separate entry were then collapsed to create the ‘final list’ of mGWAS hits.

Genes associated to mGWAS hits were determined based on the overlap of unitigs to genes. If the overlap to a gene was partial or complete, that gene was considered to be associated to the unitig. In case of no overlap, *i.e.*, when the unitig appeared completely between two genes, both the neighboring genes were considered to be linked to the unitig.

4.8. Experimental validation of top hits

The hits from the final list with a phenotype score of 3 or higher which were significant for PCN production phenotypes were selected as top hits for experimental validation. The CDS immediately upstream and downstream of each significant unitig were chosen as metabolic engineering targets to be overexpressed in the top PCN-producing strain. Any CDS which encoded ribosomal RNA was discarded from the list. If the significant unitig sequence was completely contained within a CDS, only the unitig-containing CDS was studied rather than the 2 adjacent CDS.

Each target was PCR-amplified from the genomic DNA of the strain listed on the top hits file then inserted into the backbone of plasmid pBb (RK2)1k-GFPuv using either restriction digest cloning or NEBuilder® HiFi DNA Assembly (New England Biolabs; Ipswich, MA) using the primers listed in Table S4. pBb(RK2)1k-GFPuv is a broad-range expression vector with an IPTG-inducible promoter which was gifted by Brian Pfleger at the University of Wisconsin, Madison, and used as the empty vector control (Cook et al., 2018). All plasmids used in this study are listed in Table S5.

Plasmids were transformed into the respective strain via electroporation based on the method of Choi et al. (Choi et al. 2006). Electroporations were performed by pulsing either 1.8 or 2.5 kV through a 0.1 or 0.2 cm electroporation cuvette using a MicroPulser™ Electroporation Apparatus (Bio-Rad) then recovering the reaction for 2–3 h at 30 °C. Culturing and phenazine quantification were performed as described in previous subsections.

4.9. Gene ontology enrichment analysis

To identify enriched GO-terms for significant mGWAS hits, strain DSM 21509 (highest PCN producer) was used as the reference. GO-IDs for this strain were obtained using Blast2GO v6.0.3 (Conesa et al., 2005). First, Blast2GO was used to map the annotated genome of strain DSM 21509 to proteins in the *P. chlororaphis* protein file (program: blastx; number of blast hits = 5; HSP length cutoff = 50) obtained from NCBI (Taxonomy ID: 587 753). Next, the BLAST hits were mapped to GO-identifiers from the database of the Gene Ontology Consortium (Ashburner et al., 2000; Harris et al., 2004). Lastly, GO mapped hits were annotated (Hit Filter = 2; Filter GO by taxonomy: g-proteobacteria (taxa: 1236, Gammaproteobacteria)) to obtain additional information, such as enzyme codes, enzyme names and InterPro IDs. GO-enrichment test was performed with the obtained GO-IDs using the tool GOEnrichment v2.0.1 on the Galaxy platform (Afgan et al., 2018). Annotated genes associated to significant unitigs in the final list were provided as the study set. GO annotations from the strain DSM 21509 were provided as the reference set. All enrichment tests were performed using an FDR-corrected p-value cutoff of 0.05 for enrichment.

4.10. Pathway enrichment analysis

A list of existing metabolic pathways (and corresponding genes involved) in *P. chlororaphis* strain PA23 was extracted from KEGG PATHWAY database (Kanehisa et al., 2017) (prefix: pch) and written into a custom pathway database file (.GMT format) using R 4.2.1 (RStudio, 2022.07.1). Sequences of all genes associated to significant unitigs in the final list were BLASTed against proteins of the *P. chlororaphis* strain PA23 (obtained from NCBI) using Blast2GO v6.0.3 (Conesa et al., 2005) (program: blastx; number of blast hits = 5; HSP length cutoff = 50) to find homologs. The list of PA23 gene homologs was then used as input along with the custom pathway database file to perform pathway enrichment analysis using the web version of the tool g:Profiler (Raudvere et al., 2019).

4.11. Data availability

Sequencing reads and assembled genomes for the 34 *Pseudomonas* isolates have been deposited in the NCBI SRA (BioProject ID: PRJNA932460) and NCBI GenBank databases, respectively. NCBI accession numbers for the assembled genomes have been provided in Table 1. Source data for main figures in the study has been provided in Supplementary File 7. Scripts used for collapsing mGWAS hits have been provided as Supplementary File 8.

Author contributions

ST, VT and IW conceived the project, planned the experiments, and

analyzed the data. ST collected the bacterial isolates, sequenced and assembled the whole genomes and the pangenome, and conducted the phenotype screening experiments for phenazine production and biofilm formation. EO, ST and VT performed phenotyping for growth temperature. VT implemented the mGWAS analysis and performed downstream processing of mGWAS hits. ST carried out validation experiments for the top mGWAS hits. All authors wrote and edited the manuscript.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported by the Army Research Office MURI (#W911NF1410263), Air Force Office of Scientific Research award FA9550-17-1-0270, and NSF Plants-3D 1922642.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ymben.2023.06.008>.

References

- Afgan, Enis, Baker, Dannon, Batut, Bérénice, van den Beek, Marius, Bouvier, Dave, Cech, Martin, Chilton, John, et al., 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 46 (W1), W537–W544.
- Andrews, Simon, Others, 2010. “FastQC: A Quality Control Tool for High Throughput Sequence Data.” Babraham Bioinformatics. Babraham Institute, Cambridge, United Kingdom.
- Ashburner, Michael, Ball, Catherine A., Blake, Judith A., Botstein, David, Butler, Heather, Michael Cherry, J., Davis, Allan P., et al., 2000. Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25 (1), 25–29.
- Banzhaf, Manuel, Hamish, Cl Yau, Verheul, Jolanda, Adam, Lodge, George, Kritikos, Mateus, André, Cordier, Baptiste, et al., 2020. Outer membrane lipoprotein Nlp scaffolds peptidoglycan hydrolases within multi-enzyme complexes in *Escherichia coli*. *EMBO J.* 39 (5), e102246.
- Benedetti, Ilaria, de Lorenzo, Víctor, Nikel, Pablo I., 2016. Genetic programming of catalytic *Pseudomonas putida* biofilms for boosting biodegradation of haloalkanes. *Metab. Eng.* 33 (January), 109–118.
- Bolger, Anthony M., Lohse, Marc, Usadel, Bjoern, 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15), 2114–2120.
- Brynildsrud, Ola, Bohlin, Jon, Scheffer, Lonneke, Eldholm, Vegard, 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with scary. *Genome Biol.* 17 (1), 238.
- Burr, Sarah E., Gobieli, Stefanie, Kuhnert, Peter, Goldschmidt-Clermont, Elinor, Frey, Joachim, 2010. *Pseudomonas chlororaphis* subsp. *piscium* subsp. nov., isolated from freshwater fish. *Int. J. Syst. Evol. Microbiol.* 60 (Pt 12), 2753–2757.
- Bury-Moné, Stéphanie, Nomane, Yanouira, Reymond, Nancie, Barbet, Romain, Jacquet, Eric, Imbeaud, Sandrine, Jacq, Annick, Bouloc, Philippe, 2009. Global analysis of extracytoplasmic stress signaling in *Escherichia coli*. *PLoS Genet.* 5 (9), e1000651.
- Woeng, Chin A., Thomas, F.C., Bloemberg, Guido V., van der Bij, Arjan J., van der Drift, Koen M.G.M., Schripsema, Jan, Kroon, Bernadette, Scheffer, Rudy J., et al., 1998. Biocontrol by phenazine-1-carboxamide-producing *Pseudomonas chlororaphis* PCL1391 of tomato root rot caused by *Fusarium oxysporum* F. Sp. *Radical-lycopersici*. *Mol. Plant Microbe Interact.: MPMI (Mol. Plant-Microbe Interact.)* 11 (11), 1069–1077.
- Choi, Kyoung-Hee, Kumar, Ayush, Schweizer, Herbert P., 2006. A 10-min method for preparation of highly electrocompetent *Pseudomonas aeruginosa* cells: application for DNA fragment transfer between chromosomes and plasmid transformation. *J. Microbiol. Methods* 64 (3), 391–397.
- Cho, Jae Sung, Kim, Gi Bae, Eun, Hyunmin, Moon, Cheon Woo, Sang, Yup Lee, 2022. Designing microbial cell factories for the production of chemicals. *JACS Au* 2 (8), 1781–1799.
- Clifford, Eleanor R., Bradley, Robert W., Laura, T. Wey, Lawrence, Joshua M., Chen, Xiaolong, Howe, Christopher J., Zhang, Jenny Z., 2021. Phenazines as model low-midpoint potential electron shuttles for photosynthetic bioelectrochemical systems. *Chem. Sci.* 12 (9), 3328–3338.
- Conesa, Ana, Götz, Stefan, García-Gómez, Juan Miguel, Terol, Javier, Talón, Manuel, Robles, Montserrat, 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21 (18), 3674–3676.
- Conway, H.F., Haynes, W.C., Jackson, R.W., Locke, J.M., Pridham, T.G., Sohns, V.E., Stodola, F.H., 1956. *Pseudomonas aureofaciens* kluyver and phenazine alpha-carboxylic acid, its characteristic pigment. *J. Bacteriol.* 72 (3), 412–417.
- Cook, Taylor B., Rand, Jacqueline M., Nurani, Wasti, Dylan, K. Courtney, Liu, Sophia A., Pfleger, Brian F., 2018. Genetic tools for reliable gene expression and recombining in *Pseudomonas putida*. *J. Ind. Microbiol. Biotechnol.* 45 (7), 517–527.
- De Coster, Wouter, D’Hert, Sven, Schultz, Darrin T., Cruts, Marc, Van Broeckhoven, Christine, 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34 (15), 2666–2669.
- Guo, Shuqi, Zong, Yuanna, Yue, Shengjie, Hu, Hongbo, Wang, Wei, Bilal, Muhammad, Fei, Qiang, Zhang, Xuehong, 2022. Metabolic engineering of *Pseudomonas chlororaphis* for de novo production of iodinin from glycerol. *ACS Sustain. Chem. Eng.* 10 (28), 9194–9204.
- Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., et al., 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 32 (Database issue), D258–D261.
- Haynes, W.C., Rhodes, L.J., 1962. Comparative taxonomy of crystallogenic strains of *Pseudomonas aeruginosa* and *pseudomonas* as chlororaphis. *J. Bacteriol.* 84 (5), 1080–1084.
- Hollas, Aaron, Wei, Xiaoliang, Murugesan, Vijayakumar, Nie, Zimin, Li, Bin, Reed, David, Liu, Jun, Vincent, Sprengle, Wang, Wei, 2018. A biomimetic high-capacity phenazine-based anolyte for aqueous organic redox flow batteries. *Nat. Energy* 3 (6), 508–514.
- Housseini, B Issa, Karim, Phan, Gilles, Broutin, Isabelle, 2018. Functional mechanism of the efflux pumps transcription regulators from *Pseudomonas aeruginosa* based on 3D structures. *Front. Mol. Biosci.* 5 (June), 57.
- Hubmann, Georg, Foulquié-Moreno, Maria R., Nevoigt, Elke, Duitama, Jorge, Meurens, Nicolas, Pais, Thiago M., Mathé, Lotte, et al., 2013. Quantitative trait analysis of yeast biodiversity yields novel gene tools for metabolic engineering. *Metab. Eng.* 17 (May), 68–81.
- Jaillard, Magali, Lima, Leandro, Tournoud, Maud, Mahé, Pierre, van Belkum, Alex, Lacroix, Vincent, Jacob, Laurent, 2018. A fast and agnostic method for bacterial genome-wide association studies: bridging the gap between K-mers and genetic events. *PLoS Genet.* 14 (11), e1007758.
- Kanehisa, Minoru, Furumichi, Miho, Tanabe, Mao, Sato, Yoko, Morishima, Kanae, 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45 (D1), D353–D361.
- Kaur, Gurmeet, Subramanian, Srikrishna, 2015. The ku-mar zinc finger: a segment-swapped zinc ribbon in MarR-like transcription regulators related to the ku bridge. *J. Struct. Biol.* 191 (3), 281–289.
- Kim, J., Jeon, C.O., Park, W., 2008. Dual regulation of *zwf-1* by both 2-keto-3-deoxy-6-phosphogluconate and oxidative stress in *Pseudomonas putida*. *Microbiology (Reading)* 154 (12), 3905–3916.
- King, E.O., Ward, M.K., Raney, D.E., 1954. Two simple media for the demonstration of pyocyanin and fluorescin. *J. Lab. Clin. Med.* 44 (2), 301–307.
- Kolmogorov, Mikhail, Yuan, Jeffrey, Lin, Yu, Pavel, A. Pevzner, 2019. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37 (5), 540–546.
- Kritikos, G., Banzhaf, M., Herrera-Dominguez, L., et al., 2017. A tool named Iris for versatile high-throughput phenotyping in microorganisms. *Nat. Microbiol.* 2, 17014.
- Lees, John A., Bentley, Stephen, D., 2016. Bacterial GWAS: not just gilding the lily. *Nat. Rev. Microbiol.* 14, 406.
- Lees, John A., Vehkala, Minna, Välimäki, Niko, Harris, Simon R., Chewapreecha, Claire, Croucher, Nicholas J., Martinen, Pekka, et al., 2016. Sequence element enrichment analysis to determine the genetic basis of bacterial phenotypes. *Nat. Commun.* 7 (September), 12797.
- Lemire, Joseph, Milandu, Yves, Auger, Christopher, Adam, Bignucolo, Appanna, Varun P., Appanna, Vasu D., 2010. Histidine is a source of the antioxidant, alpha-ketoglutarate, in *Pseudomonas fluorescens* challenged by oxidative stress. *FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Lett.* 309 (2), 170–177.
- Li, Ling, Li, Zhenghua, Yao, Wentao, Zhang, Xuehong, Wang, Ruiming, Li, Piwu, Yang, Kai, Wang, Tengfei, Liu, Kaiquan, 2020. Metabolic engineering of *Pseudomonas chlororaphis* qlu-1 for the enhanced production of phenazine-1-carboxamide. *J. Agric. Food Chem.* 68 (50), 14832–14840.
- Liu, Kaiquan, Hu, Hongbo, Wang, Wei, Zhang, Xuehong, 2016. Genetic engineering of *Pseudomonas chlororaphis* GP72 for the enhanced production of 2-hydroxyphenazine. *Microb. Cell Factories* 15 (1), 131.
- Liu, Wen-Hui, Yue, Sheng-Jie, Feng, Tong-Tong, Song, Li, Huang, Peng, Hu, Hong-Bo, Wang, Wei, Zhang, Xue-Hong, 2021. Characterization and engineering of LX24 with high production of 2-hydroxyphenazine. *J. Agric. Food Chem.* 69 (16), 4778–4784.
- Manni, Mosé, Matthew, R. Berkeley, Mathieu Sappey, Felipe A. Simão, Zdobnov, Evgeny M., 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* 38 (10), 4647–4654.
- Maurer, Matthew J., Lawrence, Sutardja, Pinel, Dominic, Bauer, Stefan, Muehlbauer, Amanda L., Ames, Tyler D., Skerker, Jeffrey M., Arkin, Adam P., 2017. Quantitative trait loci (QTL)-Guided metabolic engineering of a complex trait. *ACS Synth. Biol.* 6 (3), 566–581.
- Mavrodi, Dmitri V., Blankenfeldt, Wulf, Thomashow, Linda S., 2006. Phenazine compounds in fluorescent *Pseudomonas* spp. Biosynthesis and regulation. *Annu. Rev. Phytopathol.* 44, 417–445.
- Meijnen, Jean-Paul, Randazzo, Paola, María, R., Joost van den Brink, Foulquié-Moreno, Paul, Vandecruys, Stojiljkovic, Marija, Dumortier, Françoise, et al., 2016. Polygenic analysis and targeted improvement of the complex trait of high acetic acid tolerance in the yeast *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 9 (January), 5.
- Mikheenko, Alla, Pribelski, Andrey, Savelliev, Vladislav, Antipov, Dmitry, Gurevich, Alexey, 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 34 (13), i142–i150.

- Nikel, P.I., Fuhrer, T., Chavarría, M., Sanchez-Pascuala, A., Sauer, U., de Lorenzo, V., 2021. Reconfiguration of metabolic fluxes in *Pseudomonas putida* as a response to sub-lethal oxidative stress. *ISME J* 15, 1751–1766.
- O'Toole, George, A., 2011. Microtiter dish biofilm formation assay. *J. Vis. Exp.* 47 <https://doi.org/10.3791/2437> (January).
- Peix, Alvaro, Valverde, Angel, Rivas, Raúl, José, M. Igual, Ramírez-Bahena, Martha-Helena, Mateos, Pedro F., Ignacio, Santa-Regina, Rodríguez-Barrueco, Claudino, Martínez-Molina, Eustoquio, Velázquez, Encarna, 2007. Reclassification of *Pseudomonas aurantiaca* as a synonym of *Pseudomonas chlororaphis* and proposal of three subspecies, *P. Chlororaphis* subsp. *chlororaphis* subsp. nov., *P. Chlororaphis* subsp. *aureofaciens* subsp. nov., comb. Nov. And *P. Chlororaphis* subsp. *aurantiaca* subsp. nov., comb. Nov. *Int. J. Syst. Evol. Microbiol.* 57 (Pt 6), 1286–1290.
- Prjibelski, Andrey, Antipov, Dmitry, Meleshko, Dmitry, Lapidus, Alla, Korobeynikov, Anton, 2020. Using SPAdes de novo assembler. *Current Protocols in Bioinformatics/Editorial Board, Andreas D. Baxeavanis ... [et Al.* 70 (1), e102.
- Quinlan, Aaron R., Hall, Ira M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26 (6), 841–842.
- Rabaey, Korneel, Boon, Nico, Höfte, Monica, Verstraete, Willy, 2005. Microbial phenazine production enhances electron transfer in biofuel cells. *Environ. Sci. Technol.* 39 (9), 3401–3408.
- Raudvere, Uku, Kolberg, Liis, Kuzmin, Ivan, Arak, Tambet, Adler, Priit, Peterson, Hedi, Vilo, Jaak, 2019. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47 (W1), W191–W198.
- Rij, E. Tjeerd van, Wesselink, Monique, Thomas, F., Chin-A-Woeng, C., Bloembergen, Guido V., Lugtenberg, Ben J.J., 2004. Influence of environmental conditions on the production of phenazine-1-carboxamide by *Pseudomonas chlororaphis* PCL1391. *Mol. Plant Microbe Interact.: MPMI (Mol. Plant-Microbe Interact.)* 17 (5), 557–566.
- Rimsa, Vadim, Eadsforth, Thomas C., Joosten, Robbie P., Hunter, William N., 2014. High-resolution structure of the M14-type cytosolic carboxypeptidase from *Burkholderia cenocepacia* refined exploiting PDB-REDO strategies. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 70 (Pt 2), 279–289.
- Ropars, Jeanne, Didiot, Estelle, Ricardo, C., de la Vega, Rodríguez, Bennetot, Bastien, Coton, Monika, Poirier, Elisabeth, Coton, Emmanuel, Snirc, Alodie, Le Prieur, Stéphanie, Giraud, Tatiana, 2020. Domestication of the emblematic white cheese-making fungus *penicillium camemberti* and its diversification into two varieties. *Curr. Biol.: CB* 30 (22), 4441, 53.e4.
- Seemann, Torsten, 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30 (14), 2068–2069.
- Thorwall, Sarah, Schwartz, Cory, Chartron, Justin W., Wheeldon, Ian, 2020. Stress-tolerant non-conventional microbes enable next-generation chemical biosynthesis. *Nat. Chem. Biol.* 16 (2), 113–121.
- Wang, Dongping, Lee, Sung-Hee, Seeve, Candace, Yu, Jun Myoung, Leland, S., Pierson 3rd, Pierson, Elizabeth A., 2013. Roles of the *gac-rsm* pathway in the regulation of phenazine biosynthesis in *Pseudomonas chlororaphis* 30-84. *MicrobiologyOpen* 2 (3), 505–524.
- Wang, Li-Gen, Tommy, Tsan-Yuk Lam, Xu, Shuangbin, Dai, Zehan, Zhou, Lang, Feng, Tingze, Guo, Pingfan, et al., 2020. Treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. *Mol. Biol. Evol.* 37 (2), 599–603.
- Wan, Yupeng, Liu, Hongchen, Xian, Mo, Huang, Wei, 2021. Biosynthesis and metabolic engineering of 1-hydroxyphenazine in *Pseudomonas chlororaphis* H18. *Microb. Cell Factories* 20 (1), 235.
- Wick, R., Menzel, P., 2019. Filtlong: Quality Filtering Tool for Long Reads.
- Wick, R.R., 2018. Porechop: Adapter Trimmer for Oxford Nanopore Reads (Available).
- Wick, Ryan R., Judd, Louise M., Gorrie, Claire L., Holt, Kathryn E., 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* 13 (6), e1005595.
- Wongsaroj, L., Saninjak, K., Romsang, A., Duang-nkern, J., Trinachartvanit, W., Vattanaviboon, P., Mongkolsuk, S., 2018. *Pseudomonas aeruginosa* glutathione biosynthesis genes play multiple roles in stress protection, bacterial virulence and biofilm formation. *PLoS One* 13 (10), e0205815.
- Yu, Jun Myoung, Wang, Dongping, Ries, Tessa R., Pierson 3rd, Leland S., Pierson, Elizabeth A., 2018. An upstream sequence modulates phenazine production at the level of transcription and translation in the biological control strain *Pseudomonas chlororaphis* 30-84. *PLoS One* 13 (2), e0193063.
- Zhang, Manman, Kang, Zhaoqi, Guo, Xiaoting, Guo, Shiting, Xiao, Dan, Liu, Yidong, Ma, Cuiqing, Gao, Chao, Xu, Ping, 2019. Regulation of glutarate catabolism by GntR family regulator CsiR and LysR family regulator GcdR in *Pseudomonas putida* KT2440. *mBio* 10 (4). <https://doi.org/10.1128/mBio.01570-19>.
- Zhang, Xue-Xian, Chang, Hao, Tran, Sieu L., Gauntlett, Jonathan C., Cook, Gregory M., Rainey, Paul B., 2012. Variation in transport explains polymorphism of histidine and urocanate utilization in a natural *Pseudomonas* population. *Environ. Microbiol.* 14 (8), 1941–1951.
- Zhang, Xue-Xian, Rainey, Paul B., 2007. Genetic analysis of the histidine utilization (*hut*) genes in *Pseudomonas fluorescens* SBW25. *Genetics* 176 (4), 2165–2176.
- Zhou, Zheming, Charlesworth, Jane, Achtman, Mark, 2020. Accurate reconstruction of bacterial pan- and core genomes with PEPPAN. *Genome Res.* 30 (11), 1667–1679.