A Rational Design of Pseudomonas putida KT2440 capable of Anaerobic Respiration

Linde F. C. Kampers
Wageningen University & Research https://orcid.org/0000-0002-3479-3593
Jasper J. Koehorst
Wageningen University and Research
Ruben G. A. van Heck
Wageningen Universiteit en Research
Maria Suarez-Diez
Wageningen Universiteit en Research
Alfons J. M. Stams
Wageningen Universiteit en Research
Peter J. Schaap (peter.schaap@wur.nl)
https://orcid.org/0000-0002-4346-6084

Research

Keywords: Pseudomonas, Anaerobic respiration, Anaerobic fermentation, Computational design, Bioinformatics, Microbial lifestyle engineering

Posted Date: April 8th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-21150/v1

RESEARCH

A Rational Design of *Pseudomonas putida* KT2440 capable of Anaerobic Respiration

Linde FC Kampers[1], Jasper J Koehorst[1], Ruben JA van Heck[1], Maria Suarez-Diez[1], Alfons JM Stams and Peter J Schaap[1*]

*Correspondence:
peter.schaap@wur.nl
[1]Laboratory of Systems and Synthetic Biology, Wageningen University and Research Centre, Stippeneng 4, 6708 WE Wageningen, The Netherlands
Full list of author information is available at the end of the article

Abstract

Pseudomonas putida KT2440 is a metabolically versatile, HV1-certified, genetically accessible, and thus interesting microbial chassis for biotechnological applications. However, its obligate aerobic nature hampers production of oxygen sensitive products and drives up costs in large scale fermentation. The inability to perform anaerobic fermentation has been attributed to insufficient ATP production and an inability to produce pyrimidines under these conditions. Addressing these bottlenecks enabled growth under micro-oxic conditions, but does not lead to growth or survival under anoxic conditions.

Here, a data-driven approach was used to develop a rational design for a *P. putida* KT2440 derivative strain capable of anaerobic respiration. To come to the design, data derived from a genome comparison of 1628 *Pseudomonas* strains was combined with genome-scale metabolic modelling simulations and a transcriptome dataset of 47 samples representing 14 environmental conditions from the facultative anaerobe *Pseudomonas aeruginosa*.

The results indicate that the implementation of anaerobic respiration in *P. putida* KT2440 would require at least 61 additional genes of known function, at least 8 genes encoding proteins of unknown function, and 3 externally added vitamins.

Keywords: *Pseudomonas*; Anaerobic respiration; Anaerobic fermentation; Computational design; Bioinformatics; Microbial lifestyle engineering

Introduction

Pseudomonas putida KT2440 is a HV1-certified [1], genetically accessible [2, 3, 4, 5, 6, 7] and metabolically versatile [8, 9] species, which makes it an interesting adaptable industrial workhorse [10, 11, 12]. However, its strict aerobic lifestyle is an industrial disadvantage [4, 13, 14, 15, 16] as the strict requirement for dissolved O_2 results in increased costs of large-scale cultivation and may lead to unstable production rates due to inadequate local oxygen supply caused by oxygen fluctuations. Its strict aerobic nature also excludes production of O_2-sensitive enzymes, pathway intermediates or target products.

Most *Pseudomonas* species are facultative anaerobes and can use an inorganic compound such as nitrate as alternate electron receptor. This includes species that are closely related to the *P. putida* KT2440 strain, such as *P. fluorescens* and *P. denitrificans*. Only one *Pseudomonas* species is capable of anaerobic fermentation: *Pseudomonas aeruginosa* [17, 18, 19, 20]. *P. aeruginosa* is capable of arginine fermentation and pyruvate fermentation, although the latter only leads to prolonged survival under anoxic conditions, not to growth [18, 19, 20].

The relatively short evolutionary distance between *P. putida* KT2440 and facultative anaerobic *Pseudomonas* species suggests that through the implementation of a rational engineering cycle, this strain can be adapted to a facultative anaerobic lifestyle. A Design, Build, Test, Learn engineering cycle [21] was performed in earlier work [22] in an attempt to obtain an *P. putida* KT2440 strain capable of anaerobic fermentation. Using genome metabolic models (GSMs) iJP962 and iJP746 combined with a protein domain comparison (PDC) between six aerobic *Pseudomonas putida* strains including KT2440 and six facultative anaerobic *Pseudomonas* strains, three key enzymes were selected and included in the final design: acetate kinase (encoded by *ackA*), dihydroorotate dehydrogenase (*pyrK-pyrD B*) and ribonucleotide triphosphate reductase class III (*nrdD-nrdG*). This design was built and the resulting recombinant strain showed growth under micro-oxic conditions [22]. Earlier work already described an increase in survival rates upon introduction of solely acetate kinase [4, 14], and since the model predictions used in the design only considered full anoxic conditions, survival rates of the recombinant strains under anoxic conditions need to be tested.

Here, we (*i*) determined the survival rates of the recombinant strains under anoxic conditions, (*ii*) identified limitations for anaerobic growth through respiration, and (*iii*) composed a new design for a recombinant *P. putida* KT2440 capable of anaerobic respiration. In pursuit of this goal we expanded upon earlier work using the current wealth of genome data available on *P. putida* and other *Pseudomonas* species by inclusion of 1628 strains in an extensive comparison of the protein domain content [23]. Random forest, a machine learning method, was used to identify key protein domains associated with "anaerobic growth". Transcriptome data of the *Pseudomonas aeruginosa* type strain PA14 cultures grown in 14 different conditions [24] were also taken into account and integrated with previous and newly obtained GSM simulation results to arrive to a final design.

Materials and methods

Bacterial strains and cultivation conditions

Bacterial strains and plasmids are listed in in Table S1. For plasmid construction see previous work [22]. *E. coli* CC118λpir was used for cloning procedures and plasmid maintenance, and was routinely cultivated at 37°C in aerated conditions in LB medium (10 g/l tryptone, 10 g/l NaCl and 5 g/l yeast extract), optionally containing antibiotics for selection (50 μg/ml kanamycin or 50 μg/ml ampicillin as indicated). For solid medium, 15 g/l agar was added to the medium. *P. putida* KT2440 was routinely cultivated under oxic conditions at 30°C in LB medium. Experiments were performed in De Bont minimal medium (DBGA) [25] (3.88 g/l K_2HPO_4, 1.63 g/l $NaH_2PO_4 \cdot 2H_2O$, 2.00 g/l $(NH_4)_2SO_4$, 0.1 g/l $MgCl_2 \cdot 6H_2O$, 10 mg/l EDTA, 2 mg/l $ZnSO_4 \cdot 7H_2O$, 1 mg/l $CaCl_2 \cdot 2H_2O$, 5 mg/l $FeSO_4 \cdot 7H_2O$, 0.2 mg/l $Na_2MoO_4 \cdot 2H_2O$, 0.2 mg/l $CuSO_4 \cdot 5H_2O$, 0.4 mg/l $CoCl_2 \cdot 6H_2O$, 1 mg/l $MnCl_2 \cdot 2H_2O$), with 20 g/l gluconic acid as the sole carbon source. In previous work, different carbon sources were tested for optimal performance [22]. Gluconic acid was used for optimal growth by eliminating ATP consumption for substrate uptake due to passive membrane transport. The medium was supplemented with 50 μg/ml kanamycin when indicated. Precultures were prepared aerobically overnight (o/n) at 200 rpm at 30°C.

Anoxic survival experiment

Oxygen gradients served to allow the recombinant strains to grow in micro-oxic conditions as described in [22]. Anoxic cultivation of *P. putida* KT2440 recombinants unpassed or passed over oxygen gradients was performed at 30°C in 50 ml glass 20 mm aluminium crimp cap vials with rubber stoppers (Glasgerätebau Ochs Laborfachhandel e.K.) in 30 ml DeBont GA with 1 mg/l resazurin and with 50 μg/ml kanamycin as selection marker for recombinant strains. Were indicated, a 1000x diluted vitamin mix was added (0.02 g/l biotin, 0.2 g/l nicotinamide, 0.1 g/l p-aminobenzoic acid, 0.2 g/l thiamin, 0.1 g/l panthotenic acid, 0.5 g/l pyridoxamine, 0.1 g/l cyanocobalamine, 0.1 g/l riboflavine). Before inoculation, the vials were gas exchanged with CO_2/N_2. Inoculation was done with aerobically pre-cultured bacterial sample at an OD_{600} of 0.05. Approx. 8 h after inoculation, the resazurin became completely colourless, indicating full anaerobic conditions. Samples were taken using sterile CO_2 flushed 1.5" Needles (BD Microlance) and 3-5 ml syringes (ThermoFisher) to avoid O_2 exposure. Anoxic conditions were ensured as the resazurin turned from colourless to bright pink within seconds in extracted samples. Survival rates were analysed by colony forming units (CFU) determination. A dilution series was made and five drops of 10 μl per dilution were applied onto LB-agar plates without selection marker, which were incubated o/n at 30°C. Colonies were counted manually, and photos were taken of the plates. Gram-staining was performed to ensure culture purity, according to manufacturers' instructions (Gram-staining kit Machery-Nagel, Germany).

Statistical analysis

Experiments were independently repeated six times with biological triplicates in each separate experiment. Figures represent the mean values of corresponding biological triplicates and the standard deviation. The level of significance of the differences when comparing results was evaluated by means of analysis of variance (ANOVA), with α=0.05.

Genome annotation

Information on the oxygen requirements of 16989 *Pseudomonas* strains was obtained from the Gold database [26]. Per species, extensive literature research was performed to validate their aerobicity (Data S5). 1628 Genomes of facultative anaerobic and strict anaerobic strains from the *Pseudomonas* genus were obtained from the European Nucleotide Archive repository in March 2015 [27]. All genomes were de-novo annotated in SAPP [28] using Prodigal for gene prediction (version 2.6) [29], 2010] and InterProScan version 5.4-47.0 [30] for functional annotation using Pfam [31].

Comparisons of protein domain content

The positions (start and end on the protein sequence) of the protein domains and their order in a protein when multiple domains were present, were used to identify domain architecture (i.e. combinations of protein domains). Protein domain architectures were labeled by the ordered list of Pfam identifiers as described in [32]. Protein domain architectures identified in each genome sequence were stored in a

matrix, from this a binarized domain architecture presence-absence matrix was extracted and used as an input for principal component analysis using the standard R-package prcomp and hierarchical clustering using the standard R-package hclust.

Gene persistence

The persistence of a gene in a taxonomic group or group of genomes can be defined as

$$Persistence = \frac{N(orth)}{N}$$

where $N(orth)$ is the number of genomes carrying a given ortholog and N is the number of genomes considered [23]. For the set of 1628 considered genomes. Orthologous genes were identify through identity of protein domain architectures taking into account copy number. Resulting protein domain contents were analysed through protein domain comparison (PDC).

Feature selection using random forest

The random forest classification algorithm was used to classify the genome sequences in aerobic and facultative anaerobic species with the goal to identify the domains (features) responsible for the separation in these two groups (feature selection). Three hundred randomly selected genomes from aerobic and anaerobic *Pseudomonas* species were selected to train random forest models. The process was repeated one hundred times. The resulting 100 different models were used to weigh 5831 protein domains from both aerobic and anaerobic *Pseudomonas* species. Variable selection was used to identify the most influential domains for classification in aerobic and facultative anaerobic strains, yielding 100 Gini coefficients, representing the importance of a protein domain for separation per protein domain. Gini coefficients were combined into the cumulative Gini coefficient. The resulting protein domains were separated into aerobic/anaerobic specific protein domains before further analysis.

Transcriptome data analysis

A publicly available *P. aeruginosa* transcriptome data set was retrieved from GEO database (accession number GSE55197) [24]. This dataset contains 47 samples corresponding to 14 environmental conditions, including changes in growth temperature, growth stage, osmolarity, concentration of ions in the media, and surface attachment and anaerobic respiration. For every gene the log_2 fold change of its expression values was calculated in comparing every possible conditions with anaerobic respiration. Missing or infinity values arising from genes with very low counts in some condition(s) were imputed to 0 or $\pm$ 4, according to the significance of the differential expression (False discovery rate, fdr < 0.05). Normalization, fold change computations and differential expression analysis were performed using the R package DESeq [33].

Genome-scale metabolic models

In this study we used the *P. putida* genome-scale metabolic models (GSMs) iJP962, iJN746 and iJN1411 [3, 5, 34]. iJN1411 was obtained directly from the authors [34]. GSM simulations were performed as described in [22].

Results

Insertion of *acetate kinase* in *P. putida* KT2440

Previous designs to obtain *P. putida* strains surviving anoxic conditions were conceptually based on the hypothesis that anoxic survival was prevented by a lack of energy conservation and redox balancing [4, 13, 14, 16, 15]. Expression of the acetate kinase gene from *P. aeruginosa* and *E. coli* was reported to result in extended survival under anoxic conditions [4, 14]. Expression of the acetate kinase gene (*ackA*) from *E. coli* combined with class I dihydroorotate hydrogenase (*pyrK-pyrD B*) and class III ribonucleotide triphosphate reductase (*nrdD-nrdG*) from *L. lactis* successfully led to growth under micro-oxic conditions [22].

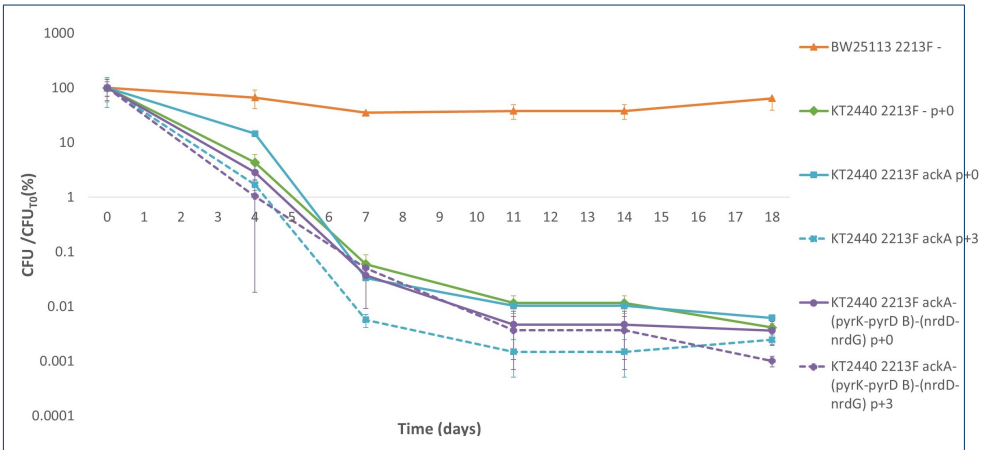

Figure 1 Anoxic survival of *P. putida* KT2440 transformed strains. Survival under anoxic conditions was determined by comparing the number of colony forming units (CFU) over time with the number of CFU at T_0. *Escherichia coli* BW25113 was used as positive control and *Pseudomonas putida* KT2440 with an empty plasmid (pS2213 -) was used as a negative control. Tested strains were *Pseudomonas putida* KT2440 with acetate kinase (pS2213 ackA) unpassed (p+0) or passed three consecutive times over oxygen gradients (p+3), and *Pseudomonas putida* KT2440 with acetate kinase, dihydroororotate dehydrogenase and ribonucleotide triphosphate reductase type II (pS2213 ackA-(pyrK-pyrD B)-(nrdD-nrdG) unpassed (p+0) or passed three consecutive times over oxygen gradients (p+3).

To determine the tolerance of a *P. putida* KT2440 negative control carrying an empty plasmid and the recombinant strain enriched with *ackA*, *pyrK-pyrD B* and *nrdD-nrdG* to anoxic conditions and to analyse the effect of an adaptation over oxygen gradients as performed earlier [22], an anoxic survival experiment of 18 days was performed. After inoculation at a standardized cell density under oxic conditions, cultures were incubated overnight in capped gas-exchanged vials in oxygen-depleted medium (see Materials and Methods). The survival rate was determined by performing colony forming unit (CFU) counts at set time points over a period of 18 days, with T_0 being the start of the experiment in anoxic conditions (Figure 1, supplementary Figures S1, S2, S3, S4). The results showed that in anoxic conditions there is no significant difference in survival rates between the negative control and any of the recombinant strains tested (ANOVA $\alpha = 0.05$). Under these conditions, only the positive control, *E. coli BW25113* harbouring an empty plasmid, survived.

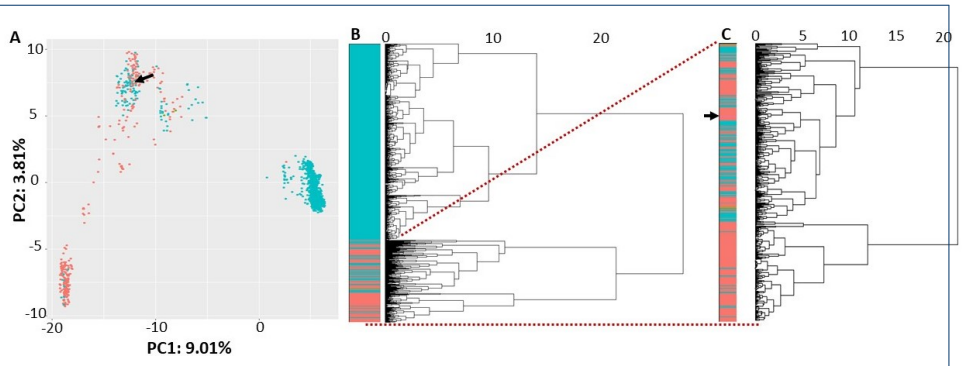

Figure 2 Protein domain content of 344 aerobic and 1284 facultative anaerobic *Pseudomonas* strains. Facultative anaerobic strains capable of respiration are indicated in blue, aerobic strains in red. (A) 2D Plot of PCA. Position of (*P. putida* KT2440 is marked with an arrow. Labels on the axes indicate fraction of the total variance explained by each component. (B) Observed distance tree based on presence/absence of protein domains. (C) Details of the main branch harbouring *P. putida* KT2440 (position indicated with an arrow). This branch consists of 138 anaerobic and 87 aerobic *Pseudomonas* species.

Design requirements for a *P. putida* KT2440 derivative strain capable of anaerobic respiration

The failure of the previous, fermentative, design [22] to grow under anoxic conditions could be explained by the heavy reliance on the two state of art genome-scale models (GSMs) used in this design, which currently do not include an accurate representation of the complete redox balance and its intricate involvement in the metabolism. Additionally, while the protein domain comparison performed in this study showed apparent differences between aerobic and anaerobic strains in availability of protein domains, this analysis was performed on a limited set of strains.

Many facultative anaerobic *Pseudomonas* species are incapable of anaerobic fermentation, but rather perform anaerobic respiration. The close phylogenetic distances between some of these facultative anaerobic *Pseudomonas* species and *P. putida* KT2440 may suggest that acquiring a facultative anaerobic lifestyle via anaerobic respiration would require less genetic changes. To come to a rational design of *P. putida* KT2440 capable of anaerobic respiration, the previous methods were thus expanded upon by (*i*) using significantly more facultative anaerobic and aerobic *Pseudomonas* strains for domain analysis, (*ii*) inclusion of iJN1411, the latest metabolic reconstruction of *P. putida* KT2440 [34], and (*iii*) incorporation of an elaborate transcriptome analysis of anaerobic respiration of *P. aeruginosa* strains grown under anoxic conditions in comparison with 13 other otherwise aerobic growth conditions [24]. Inclusion of such transcriptome data would show gene regulation due to growth under anoxic conditions, improving the design as it complements genome based methods.

For protein domain comparisons, the Pfam domain content of *P. putida* KT2440 was compared with 1627 other *Pseudomonas* strains with fully sequenced genomes. For each strain a literature search was performed to determine oxygen requirements, yielding 344 obligate aerobic strains including KT2440 and 1284 facultative anaerobic strains. Strain specific differences in protein domain content were visualized using principal component analysis (PCA), and hierarchical clustering using

domain presence/absence as input (Figure 2). Both the PCA and the hierarchical clustering show a separation between a number of the facultative anaerobic strains and the rest of the considered strains (among which *P. putida* KT2440). However, it should be noted that only a small fraction of the total variance is explained by the first two principal components. This separation is also apparent in the dendrogram, suggesting that significant differences could be found in protein domain content.

We assumed that domains essential for anaerobic respiration are highly persistent in facultative anaerobic strains, but show a lower persistence in obligate aerobic strains. The strategy to obtain this protein domain core is outlined in Figure 3. A "long list" of anaerobic protein domains was generated by comparing domain persistence between aerobic versus anaerobic strains. First a 95% persistence threshold was applied, to obtain a "domain core" of domains present in at least 95% of the genomes of "aerobic" strains and in the "anaerobic" strains analysed. These aerobic and anaerobic domain cores were used as input for subsequent comparative analysis and for the first list split into "shared between aerobic and anaerobic species" (Shared domain core), "specific for aerobic species" (Aerobe specific domain core) and "specific for anaerobic species" (Anaerobe specific domain core creating a long list of 427 anaerobe specific protein domains. A second long list was created by the same input but searching for the reverse, a separation based on domains with a very low persistency in aerobic or anaerobic strains. For this a no more than 1% threshold was applied creating the second long list of 167 anaerobe specific protein domains.

The dendrogram presented in Figure 2 indicated a possible early branch split between a large group of exclusively anaerobic *Pseudomonas* strains and a mixed group, including *P. putida* KT2440, containing 138 facultative anaerobic and 87 obligatory aerobic *Pseudomonas* strains (Figure 2 panel C). Using this split two "restricted" lists were built by comparing domain persistence as outlined above but now evaluating only *Pseudomonas* strains present in the mixed branch. For the restricted lists, a 90% threshold, and a 1% persistence threshold were used creating two anaerobic species specific protein domains lists of 170 and 248, respectively. The four different lists enriched in protein domains essential for anaerobic growth were compared to each other and manually further annotated. Results are summarized in Table 1 and Figure 3.

As outlined in the Materials and Methods section, the domain content of the facultative anaerobic and the obligatory aerobic *Pseudomonas* strains were used to train a random forest classifier with the goal to identify those domains (features) that are mostly responsible for the classification. Gini coefficients and cumulative Gini coefficients for each domain are provided in Data S9. From the 5831 domains that were used as input for the classifier, 5 were seen to have a cumulative Gini coefficient $\geq$100, as summarized in Table 1. Gini scores were added as weight to the four protein domain list derived above.

Transcriptome data obtained from *P. aeruginosa PO1*4 grown under 14 different environmental conditions including anoxic conditions [24] was re-analyzed for genes that were consistently differentially expressed during anaerobic respiration (see the Materials and Methods section for details). By calculating for every gene the log2fold change of its expression values in every possible conditions compared

with anaerobic respiration, 175 protein domains were identified. A heatmap was used to visualize up- and down-regulated genes under anoxic conditions. Regulation due to anoxic growth was considered to be significant when the same behaviour (up- or downregulation) was observed in at least 7 of the 13 pair-wise comparisons and a fold change of at least 4 was observed in at least three of these comparisons. Protein domain architectures corresponding to the selected locus tags were identified. Based on the differential expression and similar efforts in literature [13] 22 genes encompassing 35 protein domains were selected.

Genome-scale models were used to simulate anoxic conditions. The absence of any reaction products impeding growth due to the simulated lack of oxygen were pinpointed and traced back to a list of genes that either need oxygen as a substrate or that cannot be made without oxygen present, and the resulting substrates that could thus not be produced. Genes and substrates were manually verified to be essential for growth (Table 1).

Table 1: Number of unique protein domains obtained from each separate *in silico* method

Method	**# Unique protein domains**
Genome domain comparisons	
Input aerobic domain core	1029
Input anaerobic domain core	1182
Long list, 1628 strains [95% threshold]	
Shared domain core	755
Aerobe specific domain core	274
Anaerobe specific domain core	427
Long list, 1628 strains [1% threshold]	
Shared domain core	1015
Aerobe specific domain core	14
Anaerobe specific domain core	167
Restricted list, 225 strains [90% threshold]	
Shared domain core	1012
Aerobe specific domain core	17
Anaerobe specific domain core	170
Restricted list, 225 strains [1% threshold]	
Shared domain core	934
Aerobe specific domain core	95
Anaerobe specific domain core	248
Transcriptome analyses	175
GSM simulations	18
RandomForest [input]	5831
Domains with a cumulative Gini coefficient $\geq$20	360
Domains with a cumulative Gini coefficient $\geq$100	5

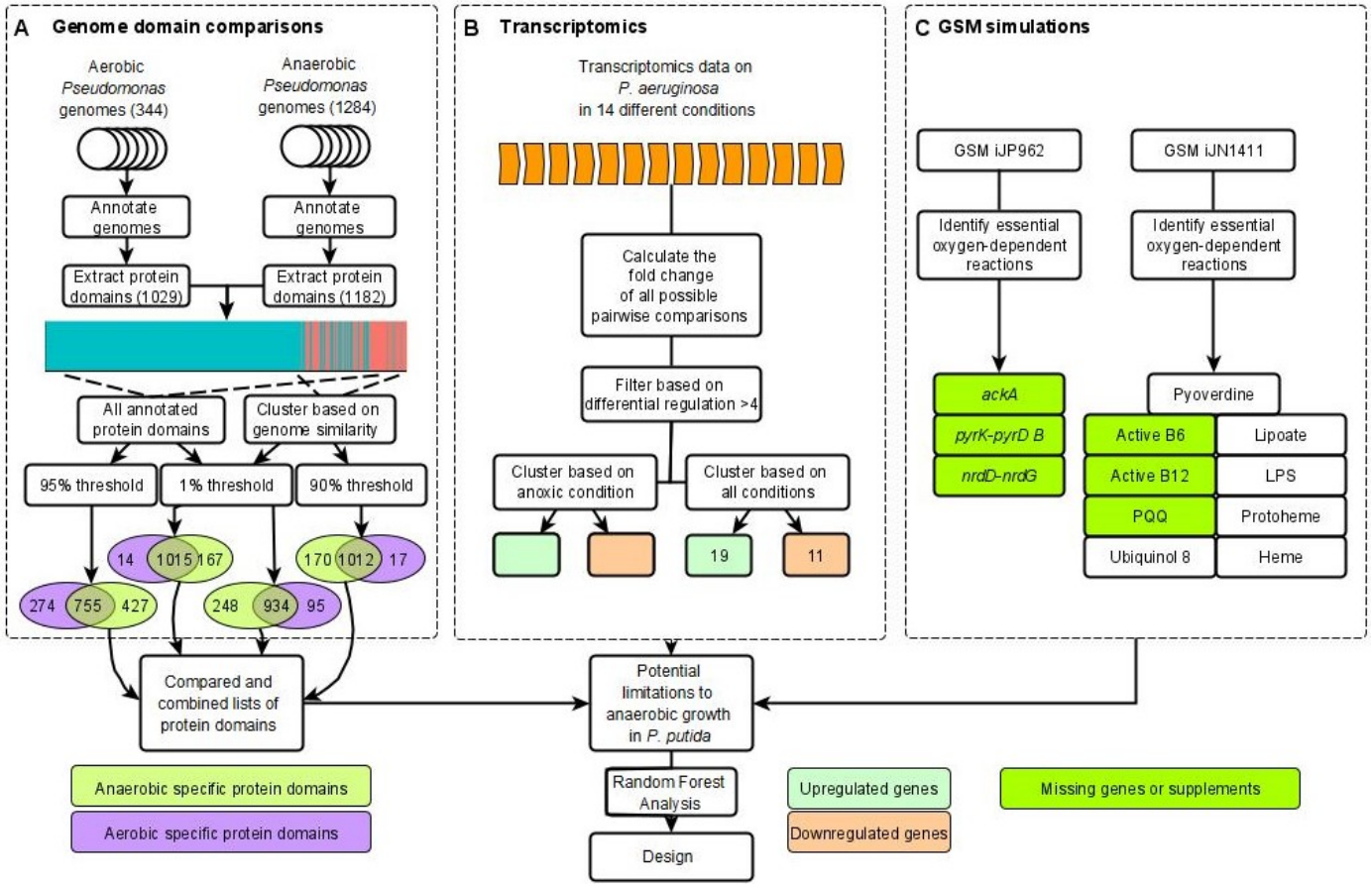

Figure 3 Overview of *in silico* approaches to identify limitations to anaerobic respiration in *P. putida*. A) Comparative genomics workflow. Genomes of the *P. putida* group and the anaerobic *Pseudomonas* group were systematically annotated using SAPP [23, 28], the protein domains were extracted, and both all domains or only the domains common to all anaerobic *Pseudomonas* species (the core domains) were selected using a 95% persistence threshold. Analysis was performed on the whole set of genomes (left) or a genome cluster of closely related strains (right). Each of these methods resulted in a list of protein domains related to an aerobic lifestyle (purple) or an anaerobic life style (light green). B) Transcriptome analysis. C) GSM simulations. GSM iJP962 [5] and iJN1411 [34] were expanded with indicated reaction sets and tested for anaerobic growth under anaerobic conditions. Colours indicate final implementation in the design (green). Model and genome base predictions were combined to obtain a final design.

Design Considerations

To obtain further insight in the requirements to build a *P. putida* KT2440 derivative strain capable of anaerobic respiration, a comparison was made between the different lists obtained (Table 1) and previous efforts [22, 4, 14, 13] resulting in an extensive overview of the many hurdles that need be overcome to build a *P. putida* KT2440 strain capable of anaerobic respiration. Lists were compared by evaluating the function of each gene starting with the encoded domain annotation, checking for domain co-existence in operonic structures, comparing metabolic functions with GSM data, and with gene regulation. The weight of each protein domain was determined using the random forest analysis (Data S9). In this way the list could be reduced to 69 genes to be included into the design and a supplement of 3 vitamins to the medium that are deemed essential for *P. putida* KT2440 to enable anaerobic respiration.

The selected genes can be separated into various categories based on their functions: Nitrogen metabolism (49 domains in 37 genes), Hydrogenases (18 domains in 16 genes), Cytochrome C (3 domains in 3 genes), Pyrimidine and amino acid biosynthesis (4 domains in 2 genes if 3 vitamins added), ATP production (3 domains in 3 genes), and Domains of Unknown Function (indirectly associated with anaerobic respiration) (8 domains).

Nitrogen metabolism

Of the 61 known genes found vital for anaerobic respiration, 37 are either directly or indirectly involved in nitrogen metabolism. With nitrate as the final electron acceptor in anaerobic respiration, compared to other final electron acceptors such as sulfate, iron(III), manganese(II), or selenate, the largest amount of energy can be conserved [35]. *P. putida* KT2440 lacks the nitrate/nitrite respiration pathway, which was resolved in earlier studies by inserting either a Nir-Nar or a Nor plasmid [13]. This resulted in extended survival under anoxic conditions, but not growth. Our transcriptomics and protein domain analysis indicated that the combination of both the Nir-Nar and the Nor operon are required (Table 2). The operons include genes required for energy conservation, cofactor biosynthesis, amino acid biosynthesis, nitrogen metabolism, nitrate-, nitrite- and nitrogen transporters, nitrate-, nitrite-, nitric oxide- and nitrous oxide reductases and several regulatory proteins (Table 2). Of the 49 protein domains or 37 genes we identified within this category, only 15 genes had been previously found (*narK1, narK2, narG, narH, narJ, narI narX, narL, nirF, nirQ, nirM, nirS, nirJ, nirL* within *nir-nar* operon, *norC, norB, norD, nosR* within the Nor operon) [13].

Previously unidentified genes in this category include many transporters and alternative mechanisms to tap indirect sources of nitrate or nitrite. *Pseudomonas* species capable of anaerobic respiration use these alternatives when nitrate or nitrite is scarce. Only genes uniformly present in species capable of anaerobic respiration and strongly associated with those of the nitrogen metabolism were considered for the design. Allantoicase (or allantoate amidinohydrolase) participates in purine metabolism, facilitating the use of purines as secondary nitrogen sources under nitrogen-limiting conditions resulting in the production of ammonia and carbon dioxide using the uricolytic pathway, which is absent in *P. putida* [36]. A second

example of an enzyme required for sourcing secondary nitrogen sources is methylaspartate ammonia-lyase. This enzyme catalyses the second step of glutamate fermentation, a process in which L-threo-3methylaspartate is converted to mesaconate and ammonia. Ureohydrolases facilitate the ammonia to urea conversion, with urea as the principle product of nitrogen excretion.

Table 2: Respiratory design of facultative anaerobic *P. putida* KT2440. Genes related to nitrogen metabolism to add for a *P. putida* KT2440 capable of anaerobic respiration

InterPro	PFAM	Name	Abbreviation	Function	Source
	PF00491	**Ureohydrolases**		Ammonia to urea conversion (principal product of nitrogen excretion)	PDC
	PF00491	Arginase	ArgI	L-arginine + H_2O ↔ L-ornithine + urea	PDC
	PF00491	Agmatinase	SpeB	Agmatine + H_2O ↔ putrescine + urea	PDC
	PF00491	Formimidoylglutaminase	HutG	N-formimidoyl-L-glutamate + H_2O ↔ L-glutamate + formamide	PDC
		Proclavaminate amidinohydrolase	Pah	Amidinoproclavaminate + H_2O ↔ proclavaminate + urea	PDC
IPR015908	PF03561			allantoate + H_2O ↔ (S)-ureidoglycolate + urea	PDC, T
IPR015868	PF04960	Glutaminase		Glutamine + H_2O → Glutamate + NH_3	PDC, T
IPR000292	PF01226	Formate/nitrite transporter		Transport of Formate/Nitrite	PDC, T
IPR025736	PF13556	PucR C-terminal helix-turn-helix domain		PucR-like transcriptional regulators	PDC, T
IPR022665	PF05034	Methylaspartate ammonia-lyase N-terminus		L-threo-3-methylaspartate → mesaconate + NH_3	PDC, T
IPR022662	PF07476	Methylaspartate ammonia-lyase C-terminus		L-threo-3-methylaspartate → mesaconate + NH_3	PDC, T
IPR000825	PF01458	**Uncharacterized protein family (UPF0051)**		Chaperone proteins for nitrogenase production	PDC, T
			NifS	Metallocluster formation nitrogenase	PDC, T
			NifU	Metallocluster formation nitrogenase	PDC, T
IPR005346	PF03658	Ubiquitin	RnfH family	Electron transport	PDC, T
	PF02508	Rnf-Nqr subunit,		Nitrogen fixation membrane protein	PDC, T
	PF03116	**Nqr2 family**		Nitrogen fixation	PDC, T
		RnfD family		Nitrogen fixation	PDC, T
		RnfE family		Nitrogen fixation	PDC, T
	PF03060	Nitronate monooxygenase		Nitrogen metabolism	PDC, T
IPR010349	PF06089	L-asparaginase II		L-asparagine + H_2O → L-aspartate + NH_3	PDC, T
PA3862	PF02423		DauB	NAD(P)H-dependent anabolic L-arginine dehydrogenase	PDC, T, [13]
PA3863	PF01266		DauA	DauBAR operon	PDC, T, [13]
PA3864	PF08348		DauR	Transcriptional regulator of the dauBAR operon	PDC, T, [13]
	PF13309				PDC, T, [13]
PA14_13750	PF07690	Nitrite extrusion protein (putative)	NarK2	Membrane proteins Transport of small molecules	PDC, T, [13]
PA14_13770	PF07690	Nitrite extrusion protein	NarK1	Membrane proteins Transport of small molecules	PDC, T, [13]
PA3875	PF14710		NarG	ATP generation	PDC, T, [13]
	PF00384				PDC, T, [13]
	PF01568				PDC, T, [13]
PA14_13800	PF13247	Nitrate reductase	NarH	β-subunit, ATP generation	PDC, T, [13]
	PF14711				PDC, T, [13]
PA14_13810	PF02613	Nitrate reductase	NarJ	λ-chain, ATP generation	PDC, T, [13]
PA14_13830	PF02665	Nitrate reductase	NarI	γ-chain, ATP generation	PDC, T, [13]
PA3878	PF02518	Two-component sensor	NarX	Nitrogen metabolism	PDC, T, [13]
	PF00672				PDC, T, [13]
	PF07730				PDC, T, [13]
	PF13675				PDC, T, [13]
PA3879	PF00072	Response regulator	NarL	Two-component response regulator	PDC, T, [13]
	PF00196				PDC, T, [13]
PA14_13850	PF04055	Heme d1 biosynthesis protein	NirJ	Heme d1 biosynthesis	PDC, T, [13]
PA0516	PF02239	Heme d1 biosynthesis protein	NirF	Heme d1 biosynthesis	PDC, T, [13]
PA0514		Heme d1 biosynthesis protein	NirL	Heme d1 biosynthesis	PDC, T, [13]
PA0520	PF07728		NirQ	Regulatory protein	PDC, T, [13]
	PF08406		CBB Q	Post-translational activation of Rubisco – photosynthesis	PDC, T, [13]
			NorQ,	Post-translational activation of Rubisco – photosynthesis	
PA14_06750	PF13442	Nitrite reductase precursor	NirM	Biosynthesis of cofactors, prosthetic groups and carriers	PDC, T, [13]
PA3870	PF00994	Molybdopterin biosynthetic protein A1	MoaA1	Biosynthesis of cofactors, prosthetic groups and carriers	PDC, [13]
	PF03453				PDC, [13]
	PF03454				PDC, [13]
PA14_13260	PF00994	Molybdopterin biosynthetic protein B1	MoaB1	Biosynthesis of cofactors, prosthetic groups and carriers	PDC
	PF00394	**Multicopper oxidase**			PDC, T
PA0519	PF13442	Nitrate reductase	NirS	ATP generation	PDC, T, [13]
		Nitrate reductase	NirS	ATP generation	PDC, T, [13]
	PF02239	Nitrate reductase	NirS	ATP generation	PDC, T, [13]
	PF05940	**NnrS protein**			PDC, T
			NirK	Reduction of nitrite to nitrous oxide	PDC, T
			Nor	Reduction of nitrite to nitrous oxide	PDC, T
PA14_06810	PF00034	Nitric-oxide reductase	NorB-NorC	Subunit B, C	PDC, T, [13]
PA14_06830	PF00115	Nitric-oxide reductase	NorB-NorC	Subunit B, C	PDC, T, [13]
PA14_06840	PF00092		NorD	Putative dinitrification protein	PDC, T, [13]
PA14_20230	PF04205		NosR	Regulatory protein for N_2O reductase	PDC, T

T, Trancriptomics; PDC, Protein Domain Comparisons; GSM, Genome Scale Modelling; [29, 30].
Printed in bold are classes of genes, the genes belonging to that class listed directly underneath.

Hydrogenases

Included in the list are 16 hydrogenases. Hydrogenases catalyse the reversible oxidation of molecular hydrogen, fulfilling a regulatory role in balancing the redox state. The redox state of the cell and the availability of O_2 are regulatory signals in facultative anaerobic species [37]. [FeFe]- And [NiFe]-hydrogenases are widely distributed under anaerobic species. These hydrogenases are only produced under anoxic conditions, and most [NiFe]-hydrogenases are inactivated by oxygen, only to be re-activated under reducing conditions [38].

Hydrogen oxidation is coupled to the reduction of electron acceptors (such as oxygen, nitrate, sulphate, carbon dioxide and fumarate). *P. putida* KT2440 lack hydrogenases necessary for the reduction of nitrogen compounds, and the necessary hydrogenase chaperones, assembly, maturation and formation proteins (Table 3).

Of the 16 proteins vital for maintaining the redox balance in anaerobic conditions only transcriptional regulator DNR has been found in previous work [13].

Table 3: Respiratory design of facultative anaerobic *P. putida* KT2440. Genes encoding for hydrogenases to add for a *P. putida* KT2440 capable of anaerobic respiration

InterPro	PFAM	Name	Abbreviation	Function	Source
IPR027394	PF14720	NiFe/NiFeSe hydrogenase small subunit C-terminal		Oxidation of molecular hydrogen	PDC
IPR001501	PF00374	Nickel-dependent hydrogenase		activation of hydrogen	PDC
IPR006894	PF04809	Hydrogenase expression protein	HupH	Hydrogenase synthesis, C-terminal conserved region	PDC
IPR000671	PF01750	Hydrogenase maturation protease		Hydrogenase maturation	PDC
IPR002780	PF01924	Hydrogenase formation hypA family	HypA-HypF	Hydrogenase formation	PDC, [38]
IPR000688	PF01155	Hydrogenase/urease nickel incorporation	HypA	[Ni,Fe]-Hydrogenase and urease metallochaperone	PDC, [38]
IPR010893	PF07449	Hydrogenase-1 expression protein	HyaE	Hydrogenase assembly	PDC
IPR023994	PF11939	[NiFe]-Hydrogenase assembly chaperone	HybE	[NiFe] Hydrogenases assembly chaperones	PDC
	PF13237	4Fe-4S dicluster domain		Mediate electron transfer	PDC
IPR007038	PF04955	HupE / UreJ protein		Hydrogenase / urease accessory proteins.	PDC
	PF00027	Transcriptional regulator	DNR	Transcriptional regulators	[13]
	PF13545				[13]

T, Trancriptomics; PDC, Protein Domain Comparisons; GSM, Genome Scale Modelling; [29, 30]

Cytochrome C

Included in the list are 3 C-type cytochromes. C-type cytochromes account for a vital step in ATP bio-generation via the proton motive force (Table 4). Aerobically, the cytochrome BC1 complex requires oxygen as electron acceptor, yielding H_2O. Anaerobically, cytochrome C 551 (NirN), C 552 and cytochrome C oxidase CBB Q transfer electrons to nitrate reductase (NirS) and nitric-oxide reductase (NorB-NorC). The importance of NirN and NirC (the precursor of NirN) was demonstrated in [13] (Table 2).

In addition, cobalamin-independent methionine synthase is important. This methionine synthase is a precursor of C 551 that can be produced without using vitamin B12 (see Pyrimidine and amino acid biosynthesis, Table 5). This might be a key component for anaerobic growth, since both the protein domain analysis and the GSM iJN1411 [34] predict that, amongst other vitamins, the active form of vitamin B12 can only be bio-generated in the presence of oxygen in *P. putida* KT2440.

The PDC also indicates the need for cytochrome C 552, and for cytochrome C oxidase CBB Q and its maturation protein (Table 2, 4). The enzyme cytochrome C nitrite reductase (C 552) catalyses the six-electron reduction of nitrite to nitrogen as one of the key steps in denitrification, nitrogen is then reduced to ammonium in the nitrogen fixation pathway, where it participates in the anaerobic energy metabolism

of dissimilatory nitrate ammonification. Expression of cytochrome CBB Q oxidase allows agronomic important diazotrophs to sustain anaerobic respiration [39].

Table 4: Respiratory design of facultative anaerobic *P. putida* KT2440. Genes related to cytochrome C to add for a *P. putida* KT2440 capable of anaerobic respiration

InterPro	PFAM	Name	Abbreviation	Function	Source
IPR003321	PF02335	Cytochrome C 552		Proton motive force cytochrome C oxidase biogenesis	PDC
PA0517	PF13442	Probable C-type cytochrome precursor	NirC	Cofactor biosynthesis	PDC, T, [13]
PA0521	PF00510	Probable cytochrome C oxidase subunit		Proton motive force	PDC, [13]

T, Trancriptomics; PDC, Protein Domain Comparisons; GSM, Genome Scale Modelling; [29, 30]

Pyrimidine and amino acid biosynthesis

Included in the list are 2 genes involved in pyrimidine and amino acid synthesis, and additional bottlenecks that can be solved by adding 3 vitamins to the medium. Earlier GSM simulations with iJP962 indicated that alternate genes must be inserted for dihydroorotate dehydrogenase and ribonucleotide triphosphate reductase type II for pyrimidine and ultimately DNA and RNA biosynthesis [22]. Both the protein domain analysis and GSM simulations using the iJN1411 metabolic model predict that cobalamin (vitamin B12), pyridoxal-5-phosphate (vitamin B6) and menaquinone (vitamin K2) cannot be produced under anoxic conditions.

Crespo *et al.* showed that class II RNRs depend on adenosylcobalamin or vitamin B12 (cobalamin) to generate its radical independently of oxygen [40]. Cobalamin is a complex essential cofactor for many enzymes mediating methylation, reduction and intramolecular rearrangements, and for methionine synthase. There is a recognised distinction between aerobic and anaerobic generation of cobalamin [41, 42]. The routes differ in terms of cobalt chelation (via CobNST complex in the aerobic pathway, via precorrin-2 with CbiK in the anaerobic pathway) and oxygen requirements. The enzymes CobI, CobG, CobJ, CobM, CobF, CobK, CobL, CobH, CobB and CobNST form the aerobic pathway. CbiK, CbiL, CbiH, CbiF, CbiG, CbiD, CbiJ, CbiET, CbiC and CbiA form the anaerobic route [41, 30, 43]. Surprisingly, the protein domain comparison yielded none of the enzymes of the anaerobic pathway for vitamin B12 synthesis, but instead CobT and CbtB, both described as important for the aerobic pathway [41]. According to the extensive analysis, these specific protein domains linked to these genes are not present in aerobic species analysed but only in anaerobic species. It was found that in the anaerobic bacterium *Eubacterium limosum*, CobT functions as an activator for a range of lower ligand substrates including DMB, determining cobamide diversity. The specific function of CbtB is unknown [41, 42].

Vitamin B6 is required for a wide variety of processes [44]. There are many vitamin B6-dependent proteins involved in amino acid biosynthesis, amino acid catabolism, antibacterial functions, iron metabolism, carbon metabolism, nucleotide utilization, cofactors for biotin, folate and heme, NAD biosynthesis, cell wall metabolism, tRNA modification, regulation of gene expression and biofilm formation.

Vitamin K2 is responsible for electron transport during anaerobic respiration. However, knock-out experiments in *E. coli* showed that upon loss of menaquinone and vitamin K1 only 3% of theorethical yield was obtained, but this was instantly

revived to 44% upon supplementing of vitamin K1 or vitamin K2 [45], indicating vitamin K1 can partially make up for the loss of vitamin K2.

Rather than inserting all missing genes, in a minimal design setup, these vitamins can be supplemented to the medium (indicated in Table 5 with ∗). To determine any immediate effect on growth or survival rates, vitamin supplementation through the medium was tested, monitoring performance of all recombinant strains under anoxic conditions. This was done parallel to a survival experiment without vitamin mix added. No difference in growth rates or survival rates was found (Figure S4, Figure S5, Data S10, Data S11).

Table 5: Respiratory design of facultative anaerobic *P. putida* KT2440. Genes related to pyrimidine and amino acid biosynthesis to add for a *P. putida* KT2440 capable of anaerobic respiration

InterPro	PFAM	Name	Abbreviation	Function	Source
IPR002751	PF01891	Cobalt uptake substrate-specific transmembrane region	Vitamin B12 ∗	VitB12 biosynthesis	PDC
IPR006538	PF06213	Cobalamin biosynthesis protein	CobT	VitB12 biosynthesis	PDC
	PF09489	Probable cobalt transporter subunit	CbtB	VitB12 biosynthesis	PDC
	PF10531	SLBB domain	Vitamin B12	Vit B12 uptake	PDC
		Adenosylcobalamin	Vitamin B12	Cofactor for enzymes and proteins	GSM
	PF02621	Menaquinone	Vitamin K2 ∗	Electron transport	PDC, [45]
		Pyridoxal 5 phosphate	Vitamin B6 ∗	Cofactor for proteins and enzymes	GSM, [44]
		Dihydroorotate dehydrogenase	PyrK-PyrD B	Pyrimidine biosynthesis	GSM, [22]
		Ribonucleotide triphosphate reductase type II	NrdD-NrdG	Pyrimidine biosynthesis	GSM, [22]

T, Trancriptomics; PDC, Protein Domain Comparisons; GSM, Genome Scale Modelling; ∗ can be added as vitamin to medium; [29, 30]

ATP Generation

Of the 61 genes of known function required for anaerobic respiration, 3 are involved in ATP generation. Both the protein domain analysis, transcriptomics data and metabolic modelling with iJP962 and iJN1411 indicate that ATP production remains one of the main bottlenecks to tackle. Earlier work has come to the same conclusion and tackled this by insertion of genes for acetate production or ethanol production [4, 14]. Our protein domain analysis has elucidated specific ATPases that only occur in anaerobic strains, providing an alternative to ATP production by fermentation (Table 6).

Table 6: Respiratory design of facultative anaerobic *P. putida* KT2440. Genes related to ATP generation to add for a *P. putida* KT2440 capable of anaerobic respiration

InterPro	PFAM	Name	Abbreviation	Function	Source
IPR002736	PF01874	ATP:dephospho-CoA triphosphoribosyl transferase		Triphosphoribosyl-dephospho-CoA production	GSM
IPR017557	PF10620	Phosphoribosyl-dephospho-CoA transferase	MdcG	Triphosphoribosyl-dephospho-CoA production	GSM
		Acetate kinase	AckA	ADP to ATP conversion by acetate production	PDC, GSM, [22]

T, Trancriptomics; PDC, Protein Domain Comparisons; GSM, Genome Scale Modelling; [29, 30]

Domains of Unknown Function

The protein domain analysis resulted in 270 unique protein domains of unknown function occurring in the genomes of anaerobic strains but not in aerobic strains. Based on contextual information, 8 were identified as important for anaerobic respiration. These were included in the design (Table 7). Similarly, 28 protein domains of unknown function were associated with virology factors or immunity, and could be excluded from the design. This leaves 244 protein domains of which the function is unknown and which can thus not be completely excluded from this design.

Table 7: Respiratory design of facultative anaerobic *P. putida* KT2440. Domains of unknown function to add for a *P. putida* KT2440 capable of anaerobic respiration

InterPro	PFAM	Name	Abbreviation	Function	Source
	PF09086	Domain of unknown function	DUF1924	ND	PDC
IPR013039	PF07627	Domain of unknown function	DUF1588	ND	PDC
IPR013036	PF07626	Domain of unknown function	DUF1587	ND	PDC
IPR013042	PF07631	Domain of unknown function	DUF1592	ND	PDC
IPR013043	PF07637	Domain of unknown function	DUF1595	ND	PDC
IPR011727	PF09601	Domain of unknown function	DUF2459	ND	PDC
	PF12981	Domain of unknown function	DUF3865	ND	PDC
	PF02026	Domain of unknown function	RyR domain	ND	PDC

T, Trancriptomics; PDC, Protein Domain Comparisons; GSM, Genome Scale Modelling; [29, 30]

Discussion

No extended survival under anoxic conditions after acetate kinase integration

Our previous rational design [22] was based on two genome-scale models and genome domain comparison analysis of six facultative anaerobic *Pseudomonas* species compared to six obligatory aerobic *Pseudomonas putida* species. Under micro-oxic conditions, the addition of acetate kinase, dihydroorotate dehydrogenase and class II ribonucleotide triphosphate reductase lead to growth.

In our hands there was no extended survival under anoxic conditions of the recombinant strains upon introduction of *ackA*. It is extremely challenging to acquire anoxic conditions. Both the medium and the headspace must be treated to completely remove oxygen from the start of the experiment, otherwise oxygen depletion takes up to 12h. Further, the medium must be prepared with L-cysteine or sodium thioglycollate to actively remove oxygen. Without these precautions, the medium can be very easily oxygenated. Small stopper-capped vials are preferred strongly over screw-cap vials, in which oxygen leaks frequently occurred [22]. Resazurin staining indicates when levels drop below a detectable level (determined with micro-electrode at 0.01 g/l dissolved oxygen, as seen in previous work [22]), but does not distinguish micro-oxic conditions from anoxic conditions.

The lack of improvement in survival rates can easily be explained when contemplating the novel design assembled in this research, as numerous essential factors such as an alternative electron acceptor or an anaerobically active cytochrome-C are missing.

Technical design issues

To enable an anaerobic lifestyle, previous designs included the introduction of between 3 and 24 genes in *P. putida* KT2440 genome [13, 14, 4, 22] but our *in silico* methods suggests that at least three times more genes are required. Novel methods developed specifically for integration of large operons or multiple genes like yTREX [46] allow incorporation of up to 14 genes at one time in *P. putida*.

The 69 genes in our design does not take into account the 244 unknown genes, which complicate the task even further. Without knowing their exact function, these genes cannot entirely be excluded from the design. At least eight of these are somehow associated with survival and/or growth in anoxic conditions [30]. The crucial roles that genes of unknown function might play was demonstrated by Hutchison and colleagues [47], who in their attempt to make a minimal bacterial genome, unexpectedly found 149 genes of unknown function to be essential for growth.

Many of the genes found in the design are closely linked to metal transport, including many hydrogenases and genes for pyrimidine and amino acid biosynthesis.

It should be considered that changes in oxygen availability drastically alters metal bioavailability as extensively reviewed in [48].

The new design compared to previous designs

We elucidated that for anaerobic growth both the *nir-nor* and *nar* operons are vital. There do exist *Pseudomonas* species that naturally have only one of these operons and are capable of nitrate to nitrite transformation. However, these strains respire nitrogen under oxic conditions only, andd have been shown to be incapable of growth in anoxic conditions [49, 50]. Building upon that, if *P. putida* KT2440 would be enriched with both the denitrification pathway and the nitrogen fixation pathway it could reduce nitrate or nitrite to ammonium, which can then be assimilated to organic compounds, transforming *P. putida* KT2440 in a diazotroph of agronomic importance [39].

The most prevalent anaerobic dissimilatory nitrate respiration regulator DNR is one of the key hydrogenases obtained from the protein domain comparison. In the facultative anaerobic *E. coli*, knock-out *fnr* mutants, an ortholog of *dnr*, were unable to grow under anoxic conditions. By DNA microarray technology it was shown that in *E. coli* 49% of the genes which differ in expression under anoxic and oxic conditions are regulated by FNR [37]. The two-component aerobic respiratory control system (ArcA and ArcB) controls gene transcription in *E. coli* under anoxic conditions. Mutations in this system are known to affect expression of over 30 operons. Most of these are repressed under anoxic conditions, but cytochrome C oxidase and pyruvate formate lyase are activated. In *E. coli*, ArcA and FNR are deemed essential for anaerobic activation [51]. In an anaerobic respiratory design of *P. putida* KT2440, it is debatable whether regulatory genes are required. We deem this advisable, in order to maintain optimal functionality of this strain under oxic conditions next to gaining the anaerobic respiration treat. These genes are thus included in the final design. However, the necessary fine-tuning of the expression levels of the regulatory genes would pose its own challenge.

We argue that for a lifestyle shift from a strict aerobic lifestyle in *P. putida* KT2440 to an anaerobic respirative one, all these genes are required. However, an increase of strain performance under micro-oxic conditions or prolonged survival rates under anoxic conditions significantly improves strain robustness in large scale bioreactors with fluctuating oxygen levels. Hence, each step towards an anaerobic lifestyle may substantially ease processes in large scale bioreactors. For enhanced performance under micro-oxic conditions, it was demonstrated that increasing ATP production alone through acetate production is enough [22]. For prolonged survival rates, however, these key elements include both Nir-Nar and Nor operons for denitrification and nitrogen fixation, cytochrome C 552, and external supplementation of the lacking vitamins. This conclusion is supported by previous findings that energy supply and redox balancing are the main bottlenecks in an anaerobic lifestyle [22, 4, 13, 14, 15, 16].

Conclusion

Increased ATP generation by insertion of *acetate kinase* via a plasmid does not lead to prolonged survival rates of *Pseudomonas putida* KT2440 under anoxic conditions. This proves that increased performance under micro-oxic conditions does

not guarantee prolonged survival under anoxic conditions. A *P. putida* KT2440 strain capable of anaerobic respiration would require the insertion of at least 69 genes into the genome and a supplement of 3 vitamins to the medium. The conversion of a strict aerobic species to a facultative anaerobic lifestyle by anaerobic respiration is a much more elaborate process than was thought before. Especially the function of DUFs and their role in anaerobic respiration must be researched, as it remains unknown how many of these should be added to this design.

Competing interests

The authors declare that they have no competing interests. All data generated or analysed during this study are included in this published article and its supplementary information files.

Author's contributions

Conceived and designed the experiments: LFCK/JJK/MSD Performed the anaerobic experiments: LFCK Performed the GSM simulations: RGAvH Performed Comparative genomics: JJK/LFCK Performed transcriptomics analysis: MSD/LFCK Protein annotation: LFCK Final design construction: LFCK/AJMS Work supervision: AJMS/PJS Wrote manuscript: LFCK Revised manuscript: MSD/AJMS/JJK/RGAvH/PJS

Acknowledgements

We gratefully acknowledge dr. P. I. Nikèl and dr. R. A. Weusthuis for insightful discussions regarding the metabolic insights and anaerobic experimentation in this project. We gratefully acknowledge financial support from the European Horizon 2020 projects EmPowerPutida (Project reference: 635536) and IBISBA (Project reference n°730976) and from the Wageningen University IPøOP project. The funders had no role in study design, data collection and analysis, or preparation of the manuscript. All authors have read and approved the manuscript.

Author details

[1]Laboratory of Systems and Synthetic Biology, Wageningen University and Research Centre, Stippeneng 4, 6708 WE Wageningen, The Netherlands. [2]Laboratory of Microbiology, Wageningen University and Research Centre, Stippeneng 4, 6708 WE Wageningen, The Netherlands.

References

1. Kampers, L.F.C., Volkers, R.J.M., Martins dos Santos, V.A.P.: *Pseudomonas putida* kt2440 is hv1 certified, not gras. Microbial Biotechnology **0**(0). doi:10.1111/1751-7915.13443. https://onlinelibrary.wiley.com/doi/pdf/10.1111/1751-7915.13443
2. Belda, E., van Heck, R.G.A., José Lopez-Sanchez, M., Cruveiller, S., Barbe, V., Fraser, C., Klenk, H.-P., Petersen, J., Morgat, A., Nikel, P.I., Vallenet, D., Rouy, Z., Sekowska, A., Martins dos Santos, V.A.P., de Lorenzo, V., Danchin, A., Médigue, C.: The revisited genome of Pseudomonas putida KT2440 enlightens its value as a robust metabolic chassis. Environmental Microbiology **18**(10), 3403–3424 (2016). doi:10.1111/1462-2920.13230
3. Nogales, J., Palsson, B.O., Thiele, I.: A genome-scale metabolic reconstruction of Pseudomonas putida KT2440: i JN746 as a cell factory. BMC Systems Biology **2**, 79 (2008). doi:10.1186/1752-0509-2-79
4. Sohn, S.B., Kim, T.Y., Park, J.M., Lee, S.Y.: In silico genome-scale metabolic analysis of Pseudomonas putida KT2440 for polyhydroxyalkanoate synthesis, degradation of aromatics and anaerobic survival. Biotechnology Journal **5**(7), 739–750 (2010). doi:10.1002/biot.201000124
5. Oberhardt, M.A., Puchałka, J., Martins dos Santos, V.A.P., Papin, J.A.: Reconciliation of Genome-Scale Metabolic Reconstructions for Comparative Systems Analysis. PLoS Computational Biology **7**(3) (2011). doi:10.1371/journal.pcbi.1001116
6. Puchałka, J., Oberhardt, M.A., Godinho, M., Bielecka, A., Regenhardt, D., Timmis, K.N., Papin, J.A., Martins dos Santos, V.A.P.: Genome-Scale Reconstruction and Analysis of the Pseudomonas putida KT2440 Metabolic Network Facilitates Applications in Biotechnology. PLoS Computational Biology **4**(10) (2008). doi:10.1371/journal.pcbi.1000210
7. van Heck, R.G., Ganter, M., dos Santos, V.A.M., Stelling, J.: Efficient reconstruction of predictive consensus metabolic network models. PLoS computational biology **12**(8) (2016)
8. Clarke, P.H.: The metabolic versatility of pseudomonads. Antonie van Leeuwenhoek **48**(2), 105–130 (1982). doi:10.1007/BF00405197
9. Bagdasarian, M., Lurz, R., Rückert, B., Franklin, F.C.H., Bagdasarian, M.M., Frey, J., Timmis, K.N.: Specific-purpose plasmid cloning vectors II. Broad host range, high copy number, RSF 1010-derived vectors, and a host-vector system for gene cloning in Pseudomonas. Gene **16**(1–3), 237–247 (1981). doi:10.1016/0378-1119(81)90080-9
10. Nikel, P.I., de Lorenzo, V.: Pseudomonas putida as a functional chassis for industrial biocatalysis: From native biochemistry to trans-metabolism. Metabolic Engineering (2018). doi:10.1016/j.ymben.2018.05.005
11. Avendaño, R., Chaves, N., Fuentes, P., Sánchez, E., Jiménez, J.I., Chavarría, M.: Production of selenium nanoparticles in pseudomonas putida kt2440. Scientific reports **6**, 37155 (2016)
12. Poblete-Castro, I., Becker, J., Dohnt, K., Santos, V.M.d., Wittmann, C.: Industrial biotechnology of Pseudomonas putida and related species. Applied Microbiology and Biotechnology **93**(6), 2279–2290 (2012). doi:10.1007/s00253-012-3928-0

13. Steen, A., Utkur, F.O., Borrero-de Acuña, J.M., Bunk, B., Roselius, L., Buhler, B., Jahn, D., Schobert, M.: Construction and characterization of nitrate and nitrite respiring Pseudomonas putida KT2440 strains for anoxic biotechnical applications. Journal of Biotechnology **163**(2), 155–165 (2013). doi:10.1016/j.jbiotec.2012.09.015
14. Nikel, P.I., de Lorenzo, V.: Engineering an anaerobic metabolic regime in Pseudomonas putida KT2440 for the anoxic biodegradation of 1,3-dichloroprop-1-ene. Metabolic Engineering **15**, 98–112 (2013). doi:10.1016/j.ymben.2012.09.006
15. Schmitz, S., Nies, S., Wierckx, N., Blank, L.M., Rosenbaum, M.A.: Engineering mediator-based electroactivity in the obligate aerobic bacterium Pseudomonas putida KT2440. Microbial Physiology and Metabolism **6**, 284 (2015). doi:10.3389/fmicb.2015.00284
16. Lai, B., Yu, S., Bernhardt, P.V., Rabaey, K., Virdis, B., Krömer, J.O.: Anoxic metabolism and biochemical production in Pseudomonas putida F1 driven by a bioelectrochemical system. Biotechnology for Biofuels **9**, 39 (2016). doi:10.1186/s13068-016-0452-y
17. Wu, M., Guina, T., Brittnacher, M., Nguyen, H., Eng, J., Miller, S.I.: The Pseudomonas aeruginosa Proteome during Anaerobic Growth. Journal of Bacteriology **187**(23), 8185–8190 (2005). doi:10.1128/JB.187.23.8185-8190.2005
18. Glasser, N.R., Kern, S.E., Newman, D.K.: Phenazine redox cycling enhances anaerobic survival in Pseudomonas aeruginosa by facilitating generation of ATP and a proton-motive force. Molecular Microbiology **92**(2), 399–412 (2014). doi:10.1111/mmi.12566
19. Eschbach, M., Schreiber, K., Trunk, K., Buer, J., Jahn, D., Schobert, M.: Long-Term Anaerobic Survival of the Opportunistic Pathogen Pseudomonas aeruginosa via Pyruvate Fermentation. Journal of Bacteriology **186**(14), 4596–4604 (2004). doi:10.1128/JB.186.14.4596-4604.2004
20. Trunk, K., Benkert, B., Quäck, N., Münch, R., Scheer, M., Garbe, J., Jänsch, L., Trost, M., Wehland, J., Buer, J., Jahn, M., Schobert, M., Jahn, D.: Anaerobic adaptation in Pseudomonas aeruginosa: definition of the Anr and Dnr regulons. Environmental Microbiology **12**(6), 1719–1733 (2010). doi:10.1111/j.1462-2920.2010.02252.x
21. Petzold, C.J., Chan, L.J.G., Nhan, M., Adams, P.D.: Analytics for metabolic engineering. Frontiers in bioengineering and biotechnology **3**, 135–135 (2015). doi:10.3389/fbioe.2015.00135
22. Kampers, L.F.C., van Heck, R.G.A., Donati, S., Saccenti, E., Volkers, R.J.M., Schaap, P.J., Suarez-Diez, M., Nikel, P.I., Martins dos Santos, V.A.P.: In silico-guided engineering of *Pseudomonas putida* towards growth under micro-oxic conditions. Microbial Cell Factories **18**(1), 179 (2019). doi:10.1186/s12934-019-1227-5
23. Koehorst, J.J., Dam, J.C.J.v., Heck, R.G.A.v., Saccenti, E., Santos, V.A.P.M.d., Suarez-Diez, M., Schaap, P.J.: Comparison of 432 *Pseudomonas* strains through integration of genomic, functional, metabolic and expression data. Scientific Reports **6**, 38699 (2016). doi:10.1038/srep38699
24. Dötsch, A., Schniederjans, M., Khaledi, A., Hornischer, K., Schulz, S., Bielecka, A., Eckweiler, D., Pohl, S., Häussler, S.: The Pseudomonas aeruginosa Transcriptional Landscape Is Shaped by Environmental Heterogeneity and Genetic Variation. mBio **6**(4), 00749–15 (2015). doi:10.1128/mBio.00749-15
25. Hartmans, S., Smits, J.P., van der Werf, M.J., Volkering, F., de Bont, J.A.M.: Metabolism of Styrene Oxide and 2-Phenylethanol in the Styrene-Degrading Xanthobacter Strain 124x. Applied and Environmental Microbiology **55**(11), 2850–2855 (1989)
26. Reddy, T.B., Thomas, A.D., Stamatis, D., Bertsch, J., Isbandi, M., Jansson, J., Mallajosyula, J., Pagani, I., Lobos, E.A., Kyrpides, N.C.: The genomes online database (gold) v. 5: a metadata management system based on a four level (meta) genome project classification. Nucleic acids research **43**(D1), 1099–1106 (2014)
27. Amid, C., Alako, B.T.F., Balavenkataraman Kadhirvelu, V., Burdett, T., Burgin, J., Fan, J., Harrison, P.W., Holt, S., Hussein, A., Ivanov, E., Jayathilaka, S., Kay, S., Keane, T., Leinonen, R., Liu, X., Martinez-Villacorta, J., Milano, A., Pakseresht, A., Rahman, N., Rajan, J., Reddy, K., Richards, E., Smirnov, D., Sokolov, A., Vijayaraja, S., Cochrane, G.: The European Nucleotide Archive in 2019. Nucleic Acids Research **48**(D1), 70–76 (2019). doi:10.1093/nar/gkz1063
28. Koehorst, J.J., van Dam, J.C.J., Saccenti, E., Martins dos Santos, V.A.P., Suarez-Diez, M., Schaap, P.J.: SAPP: functional genome annotation and analysis through a semantic framework using FAIR principles. Bioinformatics **34**(8), 1401–1403 (2018). doi:10.1093/bioinformatics/btx767
29. Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J.: Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11**, 119 (2010). doi:10.1186/1471-2105-11-119
30. Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., Hunter, S.: InterProScan 5: genome-scale protein function classification. Bioinformatics **30**(9), 1236–1240 (2014). doi:10.1093/bioinformatics/btu031
31. El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., *et al.*: The pfam protein families database in 2019. Nucleic acids research **47**(D1), 427–432 (2019)
32. Koehorst, J.J., Saccenti, E., Schaap, P.J., dos Santos, V.A.M., Suarez-Diez, M.: Protein domain architectures provide a fast, efficient and scalable alternative to sequence-based methods for comparative functional genomics. F1000Research **5** (2016)
33. Anders, S., Huber, W.: Differential expression of rna-seq data at the gene level–the deseq package. Heidelberg, Germany: European Molecular Biology Laboratory (EMBL) (2012)
34. Nogales, J., Gudmundsson, S., Duque, E., Ramos, J.L., Palsson, B.O.: Expanding The Computable Reactome In Pseudomonas putida Reveals Metabolic Cycles Providing Robustness. bioRxiv, 139121 (2017). doi:10.1101/139121
35. Carmona, M., Zamarro, M.T., Blázquez, B., Durante-Rodríguez, G., Juárez, J.F., Valderrama, J.A., Barragán, M.J.L., García, J.L., Díaz, E.: Anaerobic catabolism of aromatic compounds: a genetic and genomic view. Microbiology and molecular biology reviews : MMBR **73**(1), 71–133 (2009). doi:10.1128/MMBR.00021-08

36. Laboratories, K.: KEGG PATHWAY: Nitrogen metabolism - Pseudomonas fluorescens Pf0-1 (2019). https://www.kegg.jp/kegg-bin/highlight_pathway?scale=1.0&map=pfo00910&keyword=anaerobic$%$20respiration
37. Kovacs, A., Rakhely, G., Balogh, J., Maroti, G., Fülöp, A., Kovacs, K.: Anaerobic regulation of hydrogenase transcription in different bacteria (2005)
38. Peters, J.W., Schut, G.J., Boyd, E.S., Mulder, D.W., Shepard, E.M., Broderick, J.B., King, P.W., Adams, M.W.W.: [FeFe]- and [NiFe]-hydrogenase diversity, mechanism, and maturation. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research **1853**(6), 1350–1369 (2015). doi:10.1016/j.bbamcr.2014.11.021
39. Pitcher, R.S., Watmough, N.J.: The bacterial cytochrome cbb3 oxidases. Biochimica et Biophysica Acta (BBA) - Bioenergetics **1655**, 388–399 (2004). doi:10.1016/j.bbabio.2003.09.017
40. Crespo, A., Blanco-Cabra, N., Torrents, E.: Aerobic Vitamin B12 Biosynthesis Is Essential for Pseudomonas aeruginosa Class II Ribonucleotide Reductase Activity During Planktonic and Biofilm Growth. Frontiers in Microbiology **9** (2018). doi:10.3389/fmicb.2018.00986
41. Fang, H., Kang, J., Zhang, D.: Microbial production of vitamin B12: a review and future perspectives. Microbial Cell Factories **16** (2017). doi:10.1186/s12934-017-0631-y
42. Moore, S.J., Lawrence, A.D., Biedendieck, R., Deery, E., Frank, S., Howard, M.J., Rigby, S.E., Warren, M.J.: Elucidation of the anaerobic pathway for the corrin component of cobalamin (vitamin b12). Proceedings of the National Academy of Sciences **110**(37), 14906–14911 (2013)
43. Roth, J.R., Lawrence, J.G., Rubenfield, M., Kieffer-Higgins, S., Church, G.M.: Characterization of the cobalamin (vitamin B12) biosynthetic genes of Salmonella typhimurium. Journal of Bacteriology **175**(11), 3303–3316 (1993)
44. Richts, B., Rosenberg, J., Commichau, F.M.: A Survey of Pyridoxal 5-Phosphate-Dependent Proteins in the Gram-Positive Model Bacterium Bacillus subtilis. Frontiers in Molecular Biosciences **6** (2019). doi:10.3389/fmolb.2019.00032
45. Wissenbach, U., Kröger, A., Unden, G.: The specific functions of menaquinone and demethylmenaquinone in anaerobic respiration with fumarate, dimethylsulfoxide, trimethylamine N-oxide and nitrate by Escherichia coli. Archives of Microbiology **154**(1), 60–66 (1990)
46. Domröse, A., Weihmann, R., Thies, S., Jaeger, K.-E., Drepper, T., Loeschcke, A.: Rapid generation of recombinant pseudomonas putida secondary metabolite producers using ytrex. Synthetic and Systems Biotechnology **2**(4), 310–319 (2017). doi:10.1016/j.synbio.2017.11.001
47. Hutchison, C.A., Chuang, R.-Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., Gill, J., Kannan, K., Karas, B.J., Ma, L., et al.: Design and synthesis of a minimal bacterial genome. Science **351**(6280) (2016)
48. Hong Enriquez, R.P., Do, T.N.: Bioavailability of metal ions and evolutionary adaptation. Life **2**(4), 274–285 (2012)
49. Hatayama, K., Kawai, S., Shoun, H., Ueda, Y., Nakamura, A.: Pseudomonas azotifigens sp. nov., a novel nitrogen-fixing bacterium isolated from a compost pile. International Journal of Systematic and Evolutionary Microbiology **55**(4), 1539–1544 (2005)
50. Yumoto, I., Yamazaki, K., Hishinuma, M., Nodasaka, Y., Suemori, A., Nakajima, K., Inoue, N., Kawasaki, K.: Pseudomonas alcaliphila sp. nov., a novel facultatively psychrophilic alkaliphile isolated from seawater. International journal of systematic and evolutionary microbiology **51**(2), 349–355 (2001)
51. Nizam, S.A., Shimizu, K.: Effects of arca and arcb genes knockout on the metabolism in escherichia coli under anaerobic and microaerobic conditions. Biochemical Engineering Journal **42**(3), 229–236 (2008)

Additional Files

Additional file 1 — Table S1

Tables

Table 8 Bacterial strains and plasmids used in this study

.xlsx file with bacterial strains and plasmids listed, including references or sources

Additional file 2 — FigureS1

Figure 4 Survival experiment of P. putida KT2440 under anoxic conditions. The CFU determination of *Pseudomonas putida* KT2440 with an empty plasmid (pS2213 -), acetate kinase (pS2213 ackA) or acetate kinase, dihydroorororotate dehydrogenase and ribonucleotide triphosphate reductase type II (pS2213 ackA-(pyrK-pyrD B)-(nrdD-nrdG) unpassed (p+0) or passed three consecutive times over oxygen gradients (p+3) survival under anoxic conditions. The experiment was repeated independently six times. All figures share the same legend. (A) Experiment 1 (B) Experiment 2 (C) Experiment 3 (D) Experiment 4

.jpg file

Additional file 3 — FigureS2

.jpg file

Figure 5 Transcriptomics of Pseudomonas aeruginosa PA01 in 15 different conditions. (A) Heatmap of up- (green) or downregulation (red) of all genes per condition. (B) All upregulated genes per condition. (C) All downregulated genes per condition.

Figure 6 Survival experiment of P. putida KT2440 under anoxic conditions. The CFU determination of *Pseudomonas putida* KT2440 with an empty plasmid (pS2213 -), acetate kinase (pS2213 ackA) or acetate kinase, dihydroororotate dehydrogenase and ribonucleotide triphosphate reductase type II (pS2213 ackA-(pyrK-pyrD B)-(nrdD-nrdG) unpassed (p+0) or passed three consecutive times over oxygen gradients (p+3) with or without vitamin mix.

Additional file 4 — FigureS3
.jpg file

Additional file 5 — FigureS4

Figure 7 Growth experiment of P. putida KT2440 under anoxic conditions. The OD_{600} determination of *Pseudomonas putida* KT2440 with an empty plasmid (pS2213 -), acetate kinase (pS2213 ackA) or acetate kinase, dihydroororotate dehydrogenase and ribonucleotide triphosphate reductase type II (pS2213 ackA-(pyrK-pyrD B)-(nrdD-nrdG) unpassed (p+0) or passed three consecutive times over oxygen gradients (p+3) with vitamin mix.

.jpg file

Additional file 6 — DataS1
.xlsx file with anaerobic cultivation Analysis 1

Additional file 7 — DataS2
.xlsx file with anaerobic cultivation Analysis 2

Additional file 8 — DataS3
..xlsx file with anaerobic cultivation Analysis 3

Additional file 9 — DataS4
.xlsx file with anaerobic cultivation Analysis 4

Additional file 10 — DataS5
.xlsx file with *Pseudomonas* selection database as obtained via GOLDDatabase, including extra information and sources.

Additional file 11 — DataS6
.xlsx file with PDC data

Additional file 12 — DataS7
.xlsx file with Transcriptomics data

Additional file 13 — DataS8
.xlsx file with GSMsimulation data

Additional file 14 — DataS9
.xlsx file with random forest data

Additional file 15 — DataS10
.xlsx file with anaerobic cultivation Analysis 5

Additional file 16 — DataS11
.xlsx file with anaerobic cultivation Analysis 6

Figures

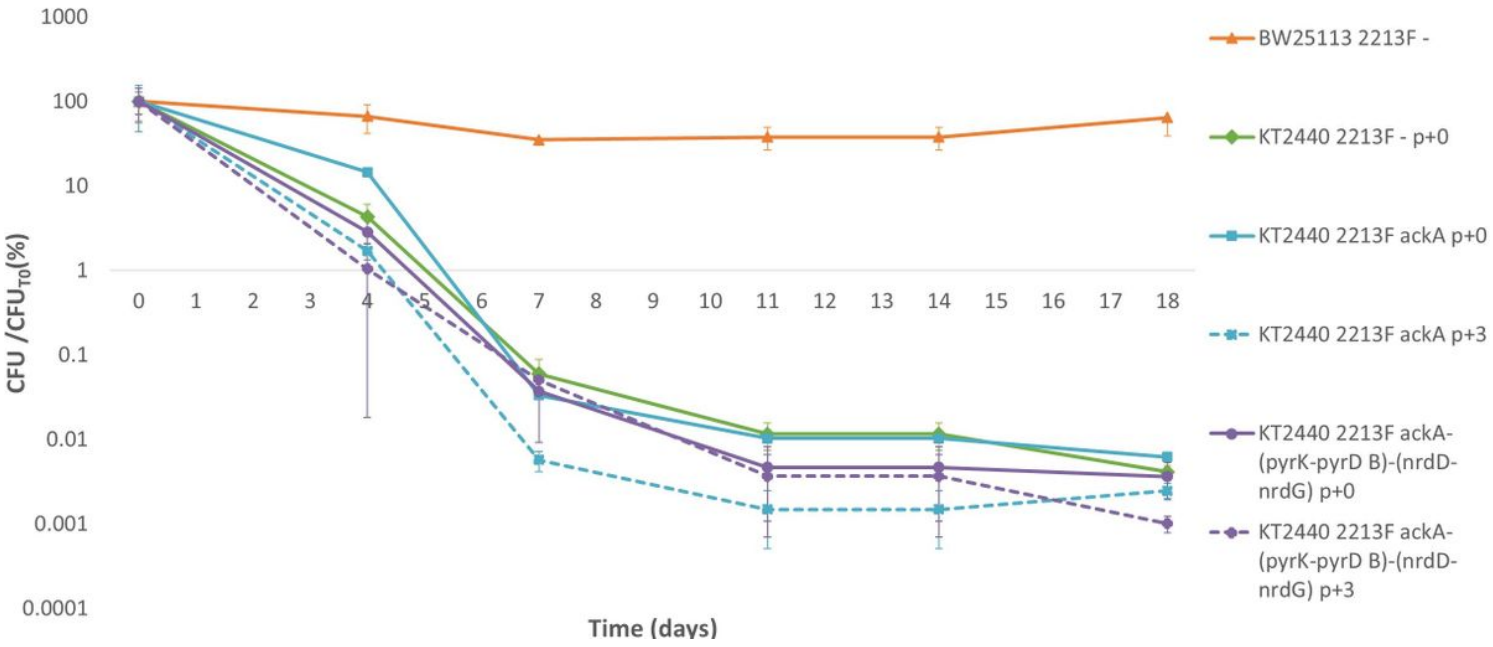

Figure 1

Anoxic survival of P. putida KT2440 transformed strains. Survival under anoxic conditions was determined by comparing the number of colony forming units (CFU) over time with the number of CFU at T0. Escherichia coli BW25113 was used as positive control and Pseudomonas putida KT2440 with an empty plasmid (pS2213 -) was used as a negative control. Tested strains were Pseudomonas putida KT2440 with acetate kinase (pS2213 ackA) unpassed (p+0) or passed three consecutive times over oxygen gradients (p+3), and Pseudomonas putida KT2440 with acetate kinase, dihydroorotate dehydrogenase and ribonucleotide triphosphate reductase type II (pS2213 ackA-(pyrK-pyrD B)-(nrdD-nrdG) unpassed (p+0) or passed three consecutive times over oxygen gradients (p+3).

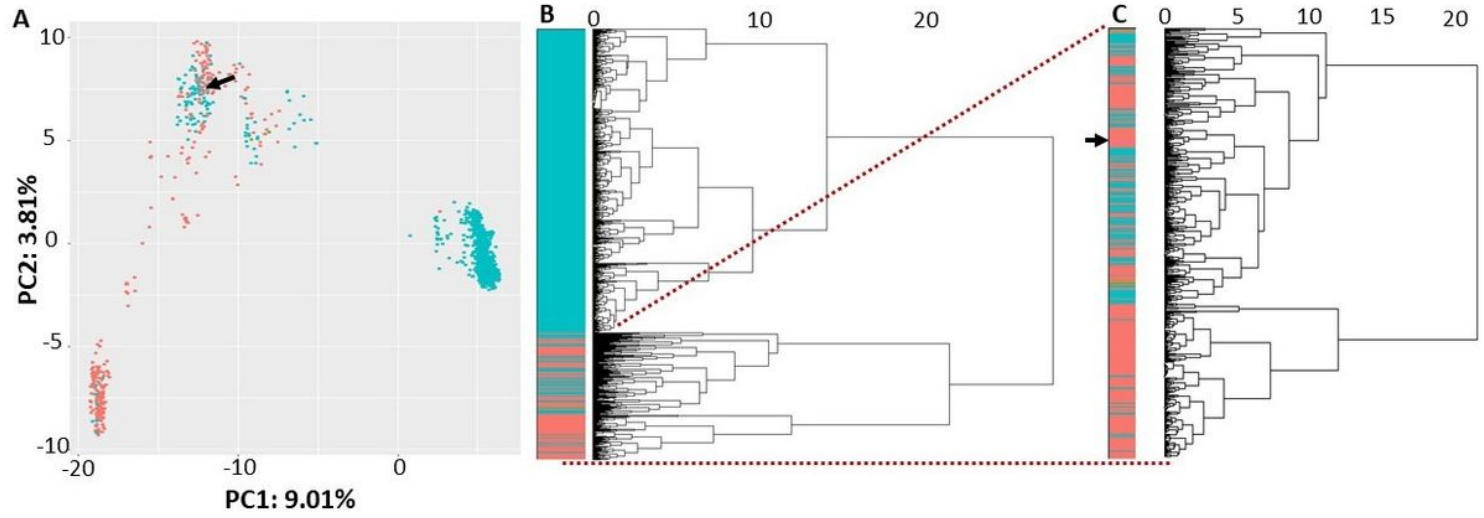

Figure 2

Protein domain content of 344 aerobic and 1284 facultative anaerobic Pseudomonas strains. Facultative anaerobic strains capable of respiration are indicated in blue, aerobic strains in red. (A) 2D Plot of PCA. Position of (P. putida KT2440 is marked with an arrow. Labels on the axes indicate fraction of the total variance explained by each component. (B) Observed distance tree based on presence/absence of protein domains. (C) Details of the main branch harbouring P. putida KT2440 (position indicated with an arrow). This branch consists of 138 anaerobic and 87 aerobic Pseudomonas species.

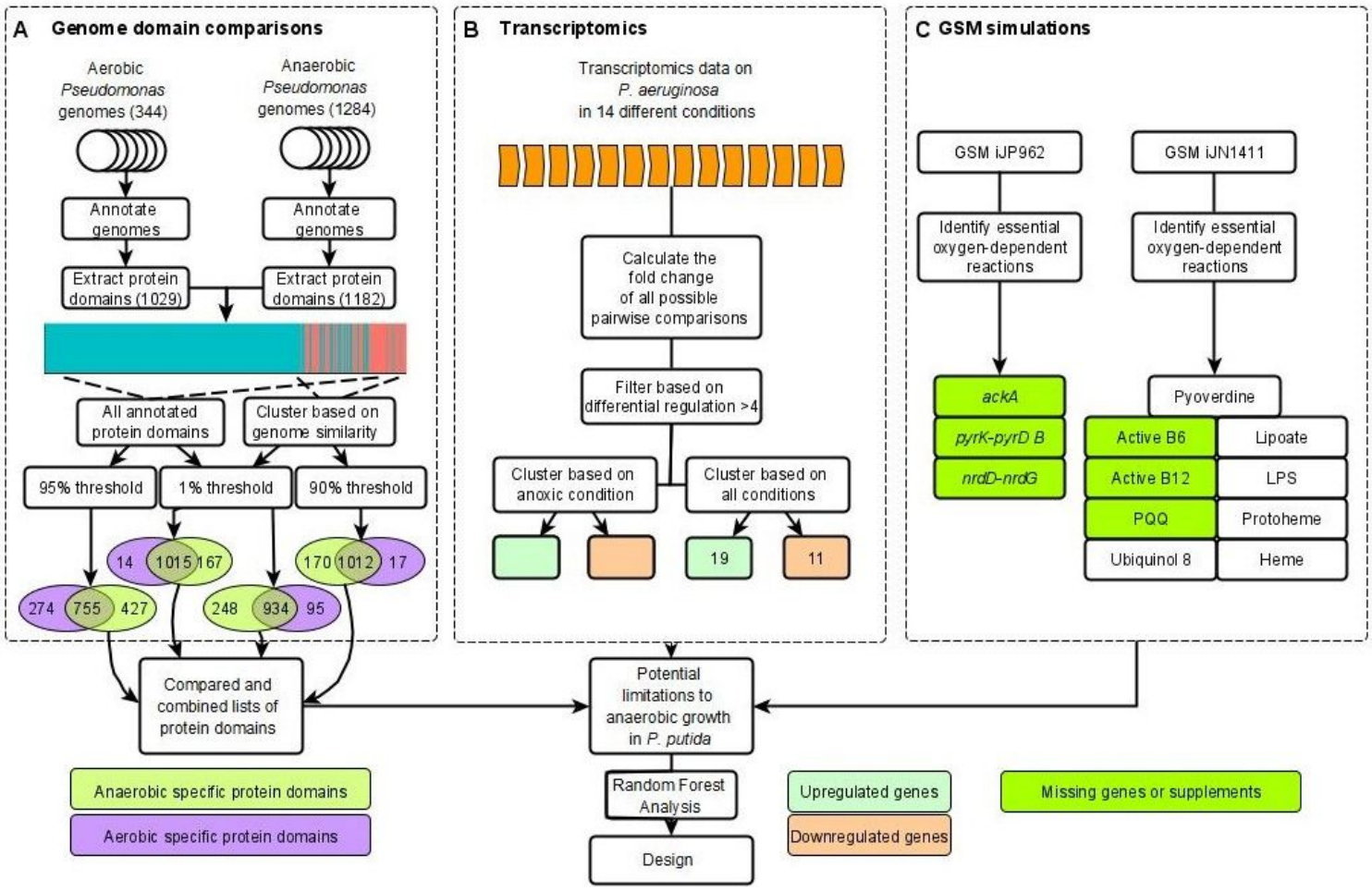

Figure 3

Overview of in silico approaches to identify limitations to anaerobic respiration in P. putida. A) Comparative genomics workflow. Genomes of the P. putida group and the anaerobic Pseudomonas group were systematically annotated using SAPP [23, 28], the protein domains were extracted, and both all domains or only the domains common to all anaerobic Pseudomonas species (the core domains) were selected using a 95% persistence threshold. Analysis was performed on the whole set of genomes (left) or a genome cluster of closely related strains (right). Each of these methods resulted in a list of protein domains related to an aerobic lifestyle (purple) or an anaerobic life style (light green). B) Transcriptome analysis. C) GSM simulations. GSM iJP962 [5] and iJN1411 [34] were expanded with indicated reaction sets and tested for anaerobic growth under anaerobic conditions. Colours indicate final implementation in the design (green). Model and genome base predictions were combined to obtain a final design.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- DataS8GSMsimulations.xlsx
- FigureS1.jpg
- DataS7Transcriptomics.xlsx
- TableS1.xlsx
- DataS9RandomForestAnalysis.xlsx
- DataS2AnaerobicSurvivalExperiment225JuneJuly2017.xlsx
- DataS10AnaerobicSurvivalExperiment52018febvitmix.xlsx
- DataS4AnaerobicSurvivalExperiment430OctoberNovember2017.xlsx
- DataS1AnaerobicSurvivalExperiment130MayJune2017.xlsx
- FigureS2.jpg
- DataS516989PseudomonasGoldDatabase.xlsx
- DataS3AnaerobicSurvivalExperiment301October2017.xlsx
- FigureS3.jpg
- FigureS4.jpg
- DataS11AnaerobicGrowthExperiment1Dec17withvitmix.xlsx
- DataS6PCA.xlsx