A pan-genome guided metabolic network reconstruction of five *Propionibacterium* species reveals extensive metabolic diversity

Tim McCubbin

Central Queensland University

R. Axayacatl Gonzalez-Garcia

Queensland Department of Education Training and Employment

Robin W. Palfreyman

University of Queensland

Chris Stowers

Corteva Inc Indianapolis

Lars K. Nielsen

University of Queensland

Esteban Marcellin (✉ e.marcellin@uq.edu.au)

University of Queensland <https://orcid.org/0000-0003-3173-7956>

Research article

Keywords: *Propionibacterium*, propionibacteria, pan-genome, genome-scale model, re-annotation, propionic acid, RAST, KBase

Posted Date: March 13th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-17101/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Genes on September 23rd, 2020. See the published version at <https://doi.org/10.3390/genes11101115>.

Abstract

Background: Propionibacteria have been studied extensively since the early 1930's due to their relevance to industry and importance as human pathogens. Still, their unique metabolism is far from fully understood. This is partly due to their signature high GC content which has hampered the acquisition of quality sequence data, the accurate annotation of the few available genomes and the functional characterization of genes. The recent completion of the genome sequences for several species enable a comparative genomic approach to annotation and provides a new opportunity to revisit our understanding of their metabolism.

Results: Using pan-genome analysis combined with the reconstruction of the first high-quality Propionibacterium genome-scale metabolic model and a pan-Propionibacterium metabolic model we demonstrate that Propionibacterium have an unexpected diversity in central carbon metabolism and a hidden layer of metabolic complexity. This combined approach gave us new insights into the evolution of Propionibacterium metabolism and led us to propose a novel, putative ferredoxin-linked energy conservation strategy. The pan-genomic approach highlighted key differences in Propionibacterium metabolism that reflect adaptation to their environment. Results were mathematically captured in genome-scale metabolic reconstructions which can be used to further explore metabolism using metabolic modelling techniques.

Conclusions: Overall, the data provides a platform to explore Propionibacterium metabolism and provides a tool for the rational design of strains.

Background

Propionibacteria are characterized by an unusually efficient anaerobic energy metabolism. This is primarily due to the Wood-Werkman cycle, a pathway that facilitates production of propionate while maximising energy capture [1, 2]. The genus is composed of two groups, the classical or dairy propionibacteria and the cutaneous. Dairy propionibacteria such as *P. acidipropionici* and *P. freudenreichii* are found in dairy products and the rumen and are associated with the ability to utilise a wide variety of carbon sources. They are used in various biotechnology applications: as starter cultures in cheese manufacture, for the production of vitamin B12 and the production of propionic acid for food applications. Some strains have the potential to produce propionic acid at high titres and yields nearing those required to replace the chemical production of industrial-grade propionic acid [3, 4]. On the other hand, cutaneous propionibacteria play an important role in secondary infections and the human metagenome and typically inhabit different regions of the skin, reflective of their metabolic specialisation. For example *P. avidum* inhabits moist, nutrient-scarce regions rich in sweat glands such as the axilla where they are likely dependent on proteolytic activity to supply amino acids as a major carbon and energy source, whereas *P. acnes* inhabits hair follicles and appears to predominantly rely on sebum degradation via lipases [5].

The prevalence of propionibacteria in the human microbiome, importance in human health and their industrial significance dictate a need for a thorough understanding of their metabolism. While many strains have been sequenced, accurate annotations are required to fully understand and ultimately rationally engineer metabolism in propionibacteria. The development of genome-scale metabolic models could offer new opportunities for studying propionibacteria, including the design of drugs or drug targets which could, in principle, be specifically targeted towards species involved in a secondary infection or acne vulgaris by studying species-specific metabolic capabilities and interactions with the human-host [6]. In addition, such models would facilitate metabolic engineering, such as to enhance yields for industrial scale production of propionate [3, 4] using knockout and overexpression vectors developed for a number of species [7–9] .

There are several challenges to producing accurate annotations for propionibacteria. Firstly, the high GC content of actinobacteria such as propionibacteria presents unique challenges to annotation, including a higher error rate and biased read coverage in genome sequencing [10]. Moreover, gene calling algorithms are susceptible to over-predicting gene length or fail to call genes [11]. Thus, automated annotations may miss many genes and inaccurately reconstruct others. Secondly, there is no biochemically well-characterized model *Propionibacterium*. Annotation by homology may therefore produce erroneous substrate predictions. For example, homologues in two strains may be mapped to different functions. Finally, the genus is metabolically diverse. For example, the second major fermentation product, acetate, is produced by various pathways in different species, as illustrated when an attempt to enhance propionate production in *P. acidipropionici* [12] failed due to the targeting of an acetate pathway present in *P. freudenreichii* subsp. *shermanii* [13] but absent in *P. acidipropionici* [14]. A good functional annotation must capture both the characteristic features of the genus such as the methylmalonyl-CoA carboxytransferase and the unique features of the specific strain. At present, semi-automated reconstruction processes and annotation pipelines such as RAST [15–17] are unable to accomplish this task, emphasised by a recent release of *Propionibacterium* models [18] which lacked a complete Wood-Werkman cycle and failed to accurately predict the different acetate production pathways.

To address these challenges, we developed a pan-genome guided metabolic network reconstruction of *Propionibacterium*. We extend the definition of the pan-genome to a genus basis as per several previous studies [19, 20] and introduce the relatively novel concept of a genus-level genome-scale metabolic reconstruction, which summarises the core and pan-metabolic functionality of the propionibacteria genus in a stoichiometric matrix, and to date has been performed in just a few studies without significant manual curation of the draft-metabolic reconstructions [21, 22]. The metabolic reconstructions for closed *Propionibacterium* species was driven by either the draft metabolic models or strain and enzyme specific information from the literature or based on genomic-context cues from the comparison of an orthologue-based pan-genomic. *P. acidipropionici* ATCC 4875 was selected for deep experimental characterisation, including growth on chemically defined media, proteomics and transcriptomics analysis and phenotype arrays, to drive further refinement of its annotation which were subsequently propagated. Core and pan-metabolic reconstructions were then generated from the individual highly curated reconstructions and utilised to compare the metabolism of five different species. Our approach demonstrates that synergistic

complementation of pan-genome and pan-GEM (genome-scale metabolic model) analyses can overcome limitations in bioinformatics tools. The analysis revealed possible evolutionary mechanism for the Wood-Werkman cycle and suggesting a putative ferredoxin-linked electron bifurcating mechanism.

Results

Exploration of core and pan-genome size

The pan-genomic analysis is an integral feature of the reconstruction process developed here as it gives preliminary estimates of the extent of diversity inherent in the *Propionibacterium* genus. Unlike previous genus level pan-genomic analyses [19, 20], only a single representative from each species was utilised to avoid clustering bias likely to arise from a greater similarity between strains. The pan-genome of the 5 species studied was composed of 6852 clusters, 876 (13%) of which composed the core *Propionibacterium* genome and 4445 (65%) were strain-specific (S1, Additional file 1). Extrapolating from this analysis using the Tettelin [6] and Willenbrock [23] algorithms, we estimated that the core genome was between 792 and 906 clusters while the pan-*Propionibacterium* genome was open, with an additional 553 new orthologous gene families identified per new species sequenced. Together, these analyses highlight significant variability between different propionibacteria species, while phylogenetic trees also demonstrate significant intra-species variability (S1, Figure 3–4, Additional file 1).

Proteomics and transcriptomics of *P. acidipropionici*

As part of the deep characterisation of *P. acidipropionici* ATCC 4875 as a reference propionibacteria strain, proteomics and transcriptomics data were gathered on glucose, sucrose and glycerol carbon sources. Each carbon source displayed key phenotypic differences; glycerol fermentation is essentially fermentative for propionate while the sugars produce substantially more acetate, succinate, pyruvate and lactate by-products, with pyruvate and acetate yields with respect to propionate halved and doubled respectively on sucrose compared to glucose (Additional file 2, S2). These differences were coupled with the differential abundance ($P < 0.05$) of transcripts and proteins for 313 genes (9% of the genome), giving insights into the functional roles of differentially expressed genes in broader metabolism. Over all conditions 3412 transcripts (~99% of the genome) were detected with an RPKM > 1 and 1085 proteins were relatively quantified, covering 100% and 63% of the core orthologous clusters previously identified respectively.

Analysis of the 'omics data in the present study is limited to assessing alternative gene calling algorithms (S8, Additional file 1), aid in the refinement of *P. acidipropionici* ATCC 4875 as a reference reconstruction and providing a deeper understanding of propionibacteria metabolism. This data is a valuable resource to guide metabolic engineering strategies and explore regulation in future studies. For example, the contrast between glycerol and sugars provides over-expression candidates to improve propionate production, several of which have already been experimentally validated [24, 25], while the fumarate reductase and NADH dehydrogenase indicate the electron transport chain may be a novel upregulation

target. This same contrast may also be informative for studying nitrogen regulation in propionibacteria with nitrogen metabolism and nitrogen respiration extensively down regulated on glycerol (Additional file 2, S2). The correlation between the transcriptome and proteome can also identify over-expression targets unlikely to be under transcriptional control, with only 36.5% of proteins having a strong correlation to their transcript ($R^2 > 0.75$) (Additional file 2, S2).

Model modifications summary

The pan-genomic matrix is insufficient to compare the metabolic functionalities of different propionibacteria due to factors such as convergent evolution, lateral gene transfer and the splitting of true orthologues due to different evolutionary distances between species. A functional pivot of this matrix can surpass some of these issues but suffers from a lack of manual curation, particularly in the context of a metabolic network. Therefore, metabolic reconstructions of five *Propionibacterium* species were generated through a novel workflow that leverages information from the pan-genomic matrix and its functional pivot (as detailed in the methods section). This included two *P. acidipropionici* strains which are the parents of high propionate producing genome-shuffled mutants [26, 27] with industrially relevance and to give a sense of intra-species variation. Individual reconstructions were combined to create a genus-level reconstruction.

The scope of the manual curation of draft reconstructions derived through KBase using this workflow is captured in Tables 1–2. On average, reconstructions were expanded in reactions and transporters by 30% (5% lacking gene assignment) and 120% (47% without gene assignment) respectively. The benefits of the pan-genomic approach are observed in Table 1 where the relative increase in information of the reconstructions of commensal species is equal to or greater than the more highly characterized dairy species. In the final reconstructions the dairy strains had no essential nutrient requirements other than the two vitamins required by all *Propionibacterium*, biotin and pantothenate (Table 3). Commensal strains had more complex requirements; requiring additional vitamins, several amino acids in the case of *P. acnes* and the nucleotide triphosphates UTP and CTP for *P. propionicum*. Reconstructions are available as an excel file with corresponding models available as SBML files and original GenBank files in Additional file 3. The *P. freudenreichii* subsp. *shermanii* model was shown to be predictive for the design of genetic engineering strategies in a recent publication [28].

Table 1 summarises key aspects of the manual curation performed of the draft reconstructions such as the addition or removal of metabolic reactions and metabolite transporters, alterations to the gene-protein-reaction associations (GPR) and the final model size and GPR coverage.

Table 2 summarises the extent of manual curation that occurred on information from the KBase database where the model was sourced from; namely corrections to the thermodynamic bounds of reactions, corrections of compounds participating in a reaction (generally preferred over generating entirely new reactions) and the additional of reactions to describe biochemistry absent from the database.

Table 3 lists the absolute minimal nutritional additives that should be added to a basic media containing salts and a source of carbon, nitrogen, phosphate and sulphur to allow growth of the listed propionibacteria, as predicted from the metabolic reconstructions.

Discussion

The metabolism of propionibacteria

The genus-level metabolic reconstruction consisted of 1223 metabolic reactions and 279 transporters, with the core reactome consisting of 761 metabolic reactions. Of the metabolic reactions, about 16% (124) occurred only in dairy species with 3% (23) occurring in all dairy propionibacteria analysed. Conversely 95 reactions (13%) were only present in cutaneous species with 16 occurring in all cutaneous species. Small differences were observed between the two *P. acidipropionici* strains, 20 metabolic reactions were present in 55737 and absent from 4875 while 16 occurred in 4875 only; however 13 of these functionalities were added on the basis of phenotype array data could not be assigned to any genes. The number of species specific metabolic functionalities was variable; 73 reactions in *P. acidipropionici* (59 gene associated), 45 in *P. propionicum*, 21 in *P. freudenreichii* subsp. *shermanii*, 15 in *P. avidum* and just 3 in *P. acnes*. To better understand how the metabolism of the 5 species analysed is conserved and variable, the genus-level metabolic reconstruction was used to compare central carbon metabolic pathways and key metabolic nodes, namely redox metabolism, the pyruvate-phosphoenolpyruvate node which splits metabolism and regulates carbon flow, the characteristic trehalose and polyphosphate metabolism and amino acid metabolism.

Central carbon metabolism

Complete pathways for glycolysis, the Wood-Werkman cycle and the pentose phosphate pathway were found in all genomes with the exception of *P. avidum* which did not have a transaldolase (2.2.1.2). Three variations of the TCA cycle lacking a glyoxalate bypass were observed which differed in how 2-oxoglutarate was dissimilated to succinate; either with substrate-level phosphorylation via the succinyl-CoA ligase (6.2.1.5) (in *P. acidipropionici*, *P. avidum* and *P. acnes*), or without via a multifunctional *menD* enzyme encoding the 2-oxoglutarate decarboxylase (4.1.1.71) and succinate semialdehyde dehydrogenase (1.2.1.16) identified in all studied propionibacteria or a complete GABA shunt confined to the dairy species. The dairy species also contained alternative central carbon metabolic pathways absent in commensal species studied. *P. freudenreichii* subsp. *shermanii* can catabolise glucose through the *Bifidobacterium* shunt which favours acetate production and may explain the lower ratio of propionate produced relative to acetate observed in this strain [28, 29]. *P. acidipropionici* possesses and transcribes a linear variation of the Wood-Werkman cycle that utilises a sodium-pumping methylmalonyl-CoA decarboxylase with a similar structure to that of *Propionigenium modestum* [30] (mean ~ 67 RPKM, Additional file 2, S2). This complex, in combination with the methylmalonyl-CoA mutase (5.4.99.2), epimerase (5.1.99.1) and the propionyl-CoA:succinate CoA-transferase (2.8.3.-), facilitates the generation

of a membrane potential through the linear dissimilation of pyruvate to propionate via succinate and potentially pre-dated the more energetically efficient Wood-Werkman cycle. Indeed, the alpha subunit of the complex and 12S subunit of the unique methylmalonyl-CoA carboxyltransferase (2.1.3.1) of propionibacteria have high identity.

The ability of *Propionibacterium* to degrade different sugars is of interest for the renewable production of propionic acid. We briefly surveyed the capability of *Propionibacterium* to degrade the constituents of hemicellulose, a major component of plants, as well as sucrose. All strains could degrade mannose and were likely capable of catabolising arabinose through the pentose phosphate pathway, although an arabinose proton symporter was only found in the dairy species. The capability to degrade xylose was only found in *P. acidipropionici* species. *P. acidipropionici* 4875 was found to contain a xylose isomerase and xylulokinase; however a xylose isomerase could not be found in strain 55737. Given we observed growth of this strain on xylose as a sole carbon source (unpublished results), promiscuous activity of the rhamnulose isomerase is likely able to compensate for this. Sucrose-specific degradative capabilities were only observed in *P. acidipropionici* and *P. propionicum*, which both contained a 6-phosphosucrose fructohydrolase, allowing catabolism of sucrose via the PTS system, as well as a sucrose-hydrolysing levanase. Both mechanisms are likely to be active in *P. acidipropionici* given upregulation of the 6-phosphosucrose fructohydrolase, fructose PTS and phosphofructokinase on sucrose (Additional file 2, S2). Less efficient catabolism of sucrose likely occurs in other strains, as observed in *P. freudenreichii* subsp. *shermanii* [31], due to promiscuous hydrolysis activities of enzymes including amylosucrase and alpha-glucosidase. Indeed, alpha-glucosidase was also upregulated in *P. acidipropionici* grown on sucrose (Additional file 2, S2). Another feature of potential industrial relevance is the *pdu* operon allowing the carboxysome-mediated degradation of 1,2-propanediol to propionate or glycerol to 3-hydroxypropionate [32] found in *P. freudenreichii* subsp. *shermanii*, in agreement with previous observations [33, 34], as well as *P. propionicum*. With this general overview of central carbon metabolism, we now focus in on the key nodes.

Acetate, ethanol and acetoin metabolism

Acetate metabolism is surprisingly variable between different propionibacteria (*Figure 2*), despite acetate being a major fermentation by-product and acetate-associated reactions act promiscuously on propionate equivalents. In fact, promiscuous activity of the propionyl-CoA:succinate CoA-transferase from the Wood-Werkman cycle is the only common enzyme allowing acetate production in the genus, suggesting other pathways have been acquired through a method such as lateral gene-transfer. All species could synthesise acetyl-CoA via the pyruvate flavodoxin oxidoreductase (PFOR) (1.2.7.1) or pyruvate dehydrogenase (1.2.4.1) while a pyruvate formate lyase (2.3.1.54) was only identified in *P. propionicum*. Further conversion to acetate with concomitant substrate-level phosphorylation was restricted to *P. freudenreichii* subsp. *shermanii* and *P. propionicum* via a phosphoacetyltransferase and acetate kinase (2.3.1.8), potentially inherited with the *pdu* operon, or the ADP-dependent acetyl-CoA synthetase (6.2.1.13) in *P. acidipropionici* which is co-located with the sodium-pumping methylmalonyl-

CoA decarboxylase. Propionibacteria lacking the *pdu* operon could directly produce acetate via a pyruvate:quinone oxidoreductase (1.2.2.2), which despite its energetic inefficiency is likely the major source of acetate in *P. acnes* and *P. avidum*, suggesting these strains are more reliant on aerobic respiration. This enzyme had a higher transcription in *P. acidipropionici* grown on glycerol where there is minimal acetate production, suggesting the role of this enzyme is likely for redox balancing, whereas the energetically beneficial acetyl-CoA synthase is upregulated on sugars (Additional file 2, S2). Notably, *P. freudenreichii* subsp. *shermanii* contains an alternative pyruvate oxidase (1.2.3.3) that consumes oxygen and produces acetyl-phosphate which likely serves as an energy yielding oxygen-tolerance mechanism, which is consistent with higher acetate production and propionate degradation when the related *P. freudenreichii* subsp. *freudenreichii* was exposed to oxygen [35].

A lack of ethanol and propanol production has been observed in *Propionibacterium* fermentations even though all strains encoded alcohol dehydrogenases (1.1.1.1) and contained proteins that can give acetaldehyde dehydrogenase (1.2.1.10) functionality. This suggests that these enzymes either are repressed or have evolved a tighter substrate specificity. Notably *P. freudenreichii* subsp. *shermanii* and *P. propionicum* have acquired a propanol dehydrogenase with the *pdu* operon [36]. Only *P. propionicum* appears to contain a true acetaldehyde dehydrogenase and has the hallmarks of an ethanol-producing strain including the pyruvate-formate lyase and would, therefore, be an attractive candidate to screen for propanol production. Acetaldehyde is also a precursor for acetoin production associated with the formation of eyes in cheese. All strains could produce acetoin aerobically via diacetyl (1.1.1.5), but utilised different paths anaerobically, either *via* acetolactate (2.2.1.6) or acetaldehyde (2.7.2.1) or both for *P. acidipropionici* (Figure 2).. *P. propionicum* was the only species containing a degradation pathway for acetoin (2.3.1.190).

Figure 2

Lactate and ferredoxin metabolism

Lactate is a key carbon source which is preferentially catabolised over glucose in some species [37] and its fermentative production plays an important role as a metabolic valve when oxidation of NADH by the respiratory chain is limiting [38, 39]. The highly endergonic production of lactate is facilitated by an NADH-dependent lactate dehydrogenase (1.1.1.27) found in all strains. Lactate oxidation is facilitated by either a menaquinone-dependent lactate dehydrogenase in *P. propionicum* or a core set of three proteins first characterised in *B. subtilis* [39] and *Shewanella* [40], which likely couples lactate oxidation to fumarate reduction via quinones or allows the direct reduction of cytochrome B, albeit weakly [41, 42]. We observed upregulation of the aforementioned complex along with a lactate permease on sugars relative to glycerol in *P. acidipropionici* 4875 where a lactate overflow was observed, in line with previous observations of transient lactate production before reconsumption during sugar fermentations in this strain [43].

Despite extensive study of lactate fermentation by propionibacteria, no attention has been given to the energetic challenge of anaerobic lactate catabolism originally highlighted in 1973 [2]. Three lactate molecules are fermented to form two propionate and one acetate molecule [2, 44] in accordance with the redox balance. However, this necessitates an energetically unfavourable reduction of NAD by a quinol; three lactate molecules are oxidised to pyruvate generating 3 quinols, two pyruvate molecules are reduced to two propionate molecules requiring two NADH and two quinols (in the absence of the NADH dehydrogenase) and the remaining pyruvate is oxidised to acetate with the generation of a quinol, NADH or ferredoxin as previously discussed (*Figure 1A*).

Given propionibacteria lack an alternative complex I permitting reverse electron transport driven by ATP hydrolysis to reduce ferredoxin or NAD with a quinone [45, 46] (*Figure 1B*), a ferredoxin based energy conservation mechanism appears to be the most likely solution. This is further supported by proteomics and transcriptomics data for *P. acidipropionici* 4875, which surprisingly suggests ferredoxin plays a major role in central carbon metabolism. Both datasets show significant upregulation of the PFOR on sugar, which appears to be the predominant source of acetyl-CoA given expression is 2–3 times higher than the pyruvate dehydrogenase (Additional file 2, S2). Conversely, expression of the pyruvate dehydrogenase is four times higher than the PFOR on glycerol illustrating that ferredoxin is associated with acetate production and not growth. Despite this, homologues to known complexes that could solve this energetic challenge could not be identified; namely an electron confurcating lactate dehydrogenase identified in *Acetobacterium woodii* [47] (*Figure 1B*) or the FixABCX complex in *Azotobacter vinelandii* which reduces ferredoxin and a quinone by oxidising 2 NADH [48].

Although no homologous genes were identified for the FixABCX complex, we postulate a similar complex exists in propionibacteria acting in the reverse direction. The flanking genes of the PFOR were identified as candidate genes for this activity based on evidence from bioinformatics analysis, 'omics data and genetic knockout. Our attention was drawn to these genes because they were co-transcribed with the PFOR (Additional file 1, S4) and were regulated identically at the protein and transcript level on different carbon sources (Additional file 2, S2). Although they were annotated as a glutamate synthase (NADPH) small chain/pyridine nucleotide-disulfide oxidoreductase and a dihydroorotate dehydrogenase in all genomes, we noted the Nfn complex catalysing the ferredoxin-mediated reduction of NADP⁺ in the presence of NADH in *Clostridium kluyveri* [49] was originally assigned the same annotations. A domain analysis revealed the glutamate synthase-like gene had a similar domain composition to *nfnB* consisting of two 4Fe–4S clusters and NADP⁺ and FAD-binding domains, but also contained an additional ferredoxin binding site, whereas the dihydroorotate dehydrogenase-like gene showed weak similarity to the *nfnA* of *C. kluyveri* and a higher similarity to type II dihydropyrimidine dehydrogenases which are directly associated with quinones. A CRISPR-mediated gene knockout in *P. freudenreichii* subsp. *shermanii* (manuscript in preparation) of the PFOR flanked by these putative genes severely crippled growth, despite other potential metabolic sources of ferredoxin and a duplicated PFOR. This suggested that this PFOR was the primary source of ferredoxin in the cell and ferredoxin plays an important role in maintaining a high growth rate. On this basis, we postulate that the two aforementioned genes and the PFOR form a complex with energy-conserving electron bifurcating properties; where ferredoxin and

NAD(P) bind to the NfnB-like subunit and the NfnA-like subunit oxidises quinols instead of ferredoxin. Further studies are warranted to confirm the existence of such a complex.

Figure 1: *A novel ferredoxin-based energy conservation mechanism explains discrepancies between the literature and observed lactate metabolism in Propionibacterium.* A- catabolic pathways for the conversion of lactate to propionate and acetate are illustrated and are most representative of the dairy propionibacteria, particularly *P. acidipropionici*, where the 2:1 ratio of propionate to acetate has been observed. The anaerobic electron transport chain is illustrated (bordered in red) and can be entered at complex II (cytochrome B) via quinols, presuming translocation of 2 protons or half an ATP gain in energy, or complex I via NADH netting an additional 2 protons pumped. The balance over lactate catabolism illustrates an overall redox imbalance exists where a quinol must be used to regenerate an NADH. B- potential strategies for overcoming this redox challenge are illustrated and are referenced to the appropriate location in metabolism in A. There was no genetic evidence for the first two mechanisms in propionibacteria, however the last step is proposed based on similarities with flanking genes of the pyruvate ferredoxin oxidoreductase and the Nfn complex of *Clostridium kluyveri* which could allow for the regeneration of NADH from a quinone.

Pyruvate metabolism

Excluding the phosphotransferase system, three separate mechanisms were identified allowing the inter-conversion of pyruvate and PEP. All *Propionibacterium* contained the pyruvate kinase (2.7.1.40) and the gluconeogenic pyrophosphate-linked pyruvate phosphate di-kinase (2.7.9.1), a role supported by the latter likely having a gluconeogenic role given up-regulation on glycerol in *P. acidipropionici* (Additional file 2, S2). *P. acidipropionici* and *P. propionicum* also contained the energetically more expensive PEP synthase (2.7.9.2), which plays a key role in gluconeogenesis in other anaerobic species.

The pyruvate transcarboxylase complex, which initiates the Wood-Werkman cycle through an ATP-independent carboxyl group transfer, is a defining aspect of propionibacteria metabolism. However three separate enzymes also allow for carbon fixation to produce oxaloacetate and may promote the production of succinate. Two separate PEP carboxykinases were identified, a GTP-dependent variant (4.1.1.32) found only in *P. propionicum* and a recently sequenced polyphosphate-dependent variant [50] found in all propionibacteria studied. Both *P. acidipropionici* and *P. propionicum* contained an additional pyruvate carboxylase (6.4.1.1). The redundancy of carbon fixation mechanisms about PEP and pyruvate coupled with a gluconeogenic malate dehydrogenase (1.1.1.39) found in all strains suggests there is complex regulation around the pyruvate node.

Respiration

All propionibacteria are able to generate ATP from the Wood-Werkman cycle via a short anaerobic electron transport chain. Electrons are transferred from NADH to intracellular fumarate via an NADH

dehydrogenase (1.6.5.3) and fumarate reductase, both of which are membrane-potential linked. Generated proton-motive force is then converted to ATP through an ATP synthase. A second copy of the fumarate reductase is thought to act as a succinate dehydrogenase [51]; a theory further supported by differential clustering of the two complexes in our pan-genomic analysis (S5, Additional file 4) and transcriptomics data for *P. acidipropionici* 4875 where only a single complex was highly expressed (Additional file 2, S2). Although it has been demonstrated that NADPH can be coupled to propionate production [28], all species were only found to have a respiration-coupled NAD(P)⁺ transhydrogenase (1.6.1.2) associated with NADP reduction, suggesting NADPH couples inefficiently via a core NADPH:quinone oxidoreductase which is upregulated when *P. acidipropionici* was grown on glycerol (Additional file 2, S2).

Propionibacteria differ significantly in their capacity to use different terminal electron acceptors. Despite the genus being viewed as microaerophilic, a cytochrome C linked oxidase, oxidoreductase and NADH dehydrogenase, allowing full aerobic respiration, were identified in three *Propionibacterium* species; while *P. freudenreichii* subsp. *shermanii* and *P. propionicum* had lost most of these subunits. The frameshift in the cytochrome C subunit I protein identified in *P. acidipropionici* 4875 [14] is not present in *P. acidipropionici* 55737 which we believe to possess full aerobic respiratory capabilities. All species contained the cytochrome *bd* oxidase (1.10.3.14) giving propionibacteria a degree of aerotolerance. Nitrate metabolism was highly variable but most extensive in *P. acidipropionici* (Figure 3)..

A number of alternative electron acceptors were identified. Nitrate metabolism was highly variable between different species but most extensive in *P. acidipropionici* (Figure 3).. All species except *P. propionicum* could reduce nitrite to nitrous oxide, while only select species contained nitrate and nitrite reductases. Similar to cytochrome C, the frameshift likely inactivating EC 1.7.1.4 in *P. acidipropionici* 4875 [14] was not present in *P. acidipropionici* 55737. Sulphate reduction to hydrogen sulphide was observed only in the dairy *Propionibacterium* strains (Figure 3).. Although a previous attempt to identify the complete pathway for sulphate reduction in *P. acidipropionici* 4875 was unsuccessful [14], we were able to postulate a complete pathway from the pan-genome matrix, consistent with the phenotype array data (S6, Additional file 5). It was found that an ambiguously annotated enzyme co-annotated with both an adenylyl-sulfate reductase functionality (1.8.4.10) and phosphoadenylyl-sulfate reductase (1.8.4.8) functionality was most similar to adenylyl-sulfate reductase and likely encoded this function. The expanded capability of *P. freudenreichii* subsp. *shermanii* and *P. acnes* to utilise alternative electron acceptors was also noted. Both were previously found to contain genes allowing the use of dimethyl sulfoxide as an electron acceptor [14]. We additionally identified homologs of the *fixABCX* genes clustered in the genome and *caiABC* homologs in both genomes; suggesting that these species can catabolise L-carnitine to gamma-butyrobetaine which additionally acts as a terminal electron acceptor.

Figure 3

Polyphosphate metabolism

Both pyrophosphate and polyphosphate metabolism are distinguishing features of *Propionibacterium* metabolism and are thought to play key roles in central carbon and energy metabolism. All propionibacteria contained both a polyphosphate kinase (2.7.4.1) and polyphosphate glucokinase (2.7.1.63) allowing excess energy to be stored as polyphosphate and used to phosphorylate glucose not transported by the phosphotransferase system, as indicated by a relative upregulation on glucose relative to sucrose in *P. acidipropionici*. This activity is solely polyphosphate dependent in all species studied except *P. propionicum* which contained an ATP-specific glucokinase. A polyphosphate kinase II was also found in *P. propionicum* and *P. acnes* which has a tendency for the production of nucleotide triphosphates from diphosphates [52]. We hypothesise this enzyme plays an essential role in the synthesis of nucleotide triphosphates in *P. propionicum* which lacks the ATP-dependent equivalent. Exopolyphosphatase (3.6.1.11) was also found in all genomes which controls polyphosphate levels and likely has a regulatory function due to the many signalling roles of polyphosphate [52, 53].

Pyrophosphate is associated with a number of reactions in central carbon metabolism. The presence of an inorganic pyrophosphatase (3.6.1.1) and relatively high expression in *P. acidipropionici* (mean ~4550 RPKM, Additional file 2, S2) suggests pyrophosphate is maintained at low levels in the cell. Reactions associated with the metabolite pyrophosphate are likely thermodynamically favourable in the direction of pyrophosphate synthesis, such as the pyrophosphate-dependent fructose 6-phosphate 1-kinase (2.7.1.90) and pyruvate-phosphate dikinase (2.7.9.1) that permit gluconeogenesis. With the exception of *P. freudenreichii* subsp. *shermanii*, all genomes contained a pyrophosphate energized proton pump which could build a proton motive force from pyrophosphate generating enzymes such as the phosphoenolpyruvate carboxytransphosphorylase (4.1.1.38) and pyruvate-phosphate dikinase; indeed 'omics data suggested coupled over-expression with the latter in *P. acidipropionici* (Additional file 2, S2).

Trehalose metabolism

Propionibacteria synthesise trehalose in response to stress [54]. Three trehalose synthesis pathways and one degradation pathway were identified, with synthesis from maltose via the trehalose synthase (5.4.99.16) being the only core functionality across all species (*Figure 3*). Additionally, the two-step synthesis of trehalose from ADP-glucose and glucose-6-phosphate was identified in all strains except *P. propionicum* while a final synthesis pathway linking glycogen metabolism to trehalose synthesis was identified in *P. acidipropionici*, *P. propionicum* and *P. avidum*. While all strains except *P. freudenreichii* subsp. *shermanii* could degrade trehalose via trehalose phosphorylase, it has been suggested that trehalose synthase plays a degradation role in this species [54]. It is worth noting that *P. acidipropionici* and *P. avidum* not only contain all trehalose metabolic functionalities of *Propionibacterium* but also contain multiple copies of several genes including trehalase (3.2.1.141) and trehalose phosphorylase (2.4.1.231), while *P. acidipropionici* additionally contains glycine betaine synthesis capability. This expanded osmoprotectant metabolism in *P. acidipropionici* suggests a potential role of osmoprotectants

in alleviating acid stress to enable the high acid producing phenotype typical of the species. These findings are supported by a trehalose over-expression study in *P. acidipropionici* 4875 [9].

Amino acid biosynthesis and catabolism

Amino acid metabolism in propionibacteria is particularly relevant to industrial applications of propionibacteria given it contributes to the formation of flavour compounds in cheese, has been reported to play key roles in stress tolerance and can influence the distribution of fermentation products. The only propiogenic amino acid catabolic pathways identified were for threonine in all species and methionine in *P. acidipropionici* 55737 only, which apart from a methionine gamma lyase was otherwise identical in amino acid metabolism to strain 4875. Branched-chain amino acids were only partially degraded to fatty acid and alcohol derivatives, consistent with experimental evidence [55] and the phenotype array data collected for *P. acidipropionici* 4875 where they could serve as nitrogen but not carbon sources. While *P. avidum* is known to possess a D-isomer 2-hydroxyacid dehydrogenase activity, a candidate gene encoding this functionality in *P. avidum* but absent in other tested strains could not be identified [56].

The ability for strains to synthesise and degrade amino acids will have an influence on the redox balance of the cell and consequently, the fermentation products. Differences in the metabolic potential for amino acid synthesis or degradation between strains were quantified using the Jaccard similarity coefficient and visualised as clustergrams (*Figure 4*). Examination of the biosynthetic pathways reveal distinct clustering of commensal and dairy bacteria. Interestingly, commensal organisms were found to have reduced diversity in their biosynthetic pathways and a wider range of degradative capabilities, particularly in comparison to *P. freudenreichii* subsp. *shermanii*. These findings likely reflect the greater reliance of proteolytic activity as a source of carbon and energy among commensal propionibacteria given their nutrient-scarce microenvironment [5]. Only *P. acnes* was detected as auxotrophic for some amino acids.

Figure 4

Propionibacteria likely contain novel amino acid catabolic pathways that may contribute to the acid tolerance response

Acid tolerance catabolic activities were also variable between species. The arginine deiminase pathway and associated antiporter were identified in *P. avidum* and *P. acnes* while the glutamate decarboxylase and associated antiporter was found in *P. freudenreichii* subsp. *shermanii*. Both *P. acidipropionici* strains contained both mechanisms without antiporters, suggesting they were less specialised to a particular acid tolerance mechanism. Although these are more traditional acid tolerance mechanisms associated with Gram positive bacteria [57], there are likely many more alkaline generating mechanisms with

phenotype array data indicating *P. acidipropionici* could utilise all 20 common amino acids as a sole nitrogen source but less than half as a carbon source, suggesting they may be very capable at the deamination of amino acids. This functionality is not captured in the current model of *P. acidipropionici* because no genes associated with these functionalities were detected, as is suggested degradation capability of phenylalanine where no homologues could be found to any known catabolic pathways.

Conclusions

We have developed six genome-scale metabolic reconstructions spanning five propionibacteria to aid *in silico* investigation into the poorly understood *Propionibacterium* metabolism. A pan-genomic guided metabolic reconstruction strategy was developed to help overcome bioinformatic challenges associated with these non-model, high GC-content organisms. The parallel curation of several related species leads to a significantly improved annotation for all reconstructions.

The central carbon metabolism was found to be highly variable between species across all the key nodes surveyed in this study. This is especially true for acetate production for which no conserved pathway existed across the five *Propionibacterium* species despite being the second major fermentation product. This variability was even observed at a strain level of analysis and was further exemplified in amino acid metabolism, where commensal organisms were found to have acquired a broader range of degradative capabilities but a reduced set of biosynthetic capabilities compared to their dairy counterparts.

By exploring the functional capacity of *Propionibacterium* metabolism, we also identified clues for the evolution of the Wood-Werkman cycle and were able to propose a novel ferredoxin-linked energy-conserving strategy and putative gene candidates. This work lays the foundations for further advancement in the understanding of *Propionibacterium* metabolism and we expect the reconstructions presented here to be instrumental in designing metabolic engineering strategies.

Methods

Strains and gene annotation

Of the 130 *Propionibacterium* genomes reported in the NCBI database at the time of performing the analysis, 21/01/2015, we focussed the analysis on the 16 closed genomes due to reduced performance of the pan-genome approach when incomplete genomes are included [58]. Although pan-genomic analyses are highly sensitive to incomplete genomes, the generation and comparison of metabolic models for functional annotations is less sensitive (S7, Additional file 1). Complete genomes of *Propionibacterium* strains were downloaded from the NCBI FTP server as GenBank files and consisted of two strains of *P. acidipropionici*, *P. freudenreichii* subsp. *shermanii* CIRM-BIA1, *P. avidum* 44067, *P. propionicum* F0230a and 11 *P. acnes* strains (Table 4). Because the generation of quality metabolic reconstructions and pan-genomic comparisons is underpinned by correct and consistent gene-annotation, the performance of alternative gene calling algorithms was evaluated in *P. acidipropionici*

4875 (S8, Additional file 1). Performance was evaluated against several metrics. Functional metrics such as differential reaction annotation and gap-filling requirements using the RAST pipeline [15–17] were initially explored but were found to be less useful given small, arbitrary shifts in gene boundaries could alter an annotation from specific to general. Instead, differences in the functional annotations were consolidated and applied to GLIMMER which performed well in the second metric; the fewest intergenic peaks in RNA sequencing data. Although intergenic peaks may exist due to non-coding RNA elements which are not annotated by the RAST pipeline, such as riboswitches, the differential number of intergenic peaks between different gene calling algorithms was assumed to indicate a failure of an algorithm to call a gene given the dense coding-sequence nature of bacterial genomes and high GC content of propionibacteria. All genomes were annotated consistently and consecutively using GLIMMER implemented in the Department of Energy Systems Biology Knowledgebase (KBASE) pipeline [59].

The following table lists the strain and accession numbers of the complete genomes utilised in this work as well as if a metabolic reconstruction was generated for the strain.

Pan-genome comparisons and matrix generation

A pan-genomic analysis underpinned the metabolic reconstruction effort by supplying extensive functional and sequence-based information (Figure 5). All clustering and pan-genomic analyses were performed on GenBank files containing gene and functional annotations from KBase using the Get_homologues package [58] which comes with a well-documented manual. Prior to clustering, exceptionally long annotations, which can arise through non-specific RAST annotations, were truncated as these become the file names of clusters generated by Get_homologues and cause the program to error when they exceed the maximum number of bytes for a Linux operating system. The analyses were performed on translated protein sequences with the OrthoMCL algorithm (OMCL) [70] with a granularity parameter of 1.5, minimum coverage of 75% and maximum E-value of 10^{-5} . Given there were multiple closed genomes for two strains, the analysis was split into two stages. First an intra-species analysis was performed and representative strains for each species were selected for an inter-species analysis (Figure 5). Such an approach avoids generating clustering bias; likely to arise from the intra-species variability exceeding the inter-species variability. The availability of additional experimental data made *P. acidipropionici* ATCC 4875 the ideal representative strain for *P. acidipropionici* while *P. acnes* 6609 was chosen as the intra-species analysis suggested it was the most functionally robust and therefore most representative of the full metabolic functionality of *P. acnes*. Estimation of the core and pan-genome size was performed using the OMCL clusters with the random sampling implementation of the Tettelin [6] and Willenbrock [23] algorithms. Trees generated from Get_homologues were visualised using FigTree (available at <http://tree.bio.ed.ac.uk/software/figtree/>).

While orthologous clusters are good indicators of conserved functionality of a cluster and phylogenetic relationships, they were found to be poor representatives of functional capacity of the genus. Evidently with a small set of strains of differing evolutionary distance, the algorithms will cluster true homologues

in separate groups based on small sequence differences. For example, robust analyses using the intersecting clusters of the OMCL, BDBH and COG algorithms [58] (Figure 5) excluded core metabolic clusters including functionalities from the Wood-Werkman cycle, the pentose phosphate pathway and the gluconeogenic malic enzyme (S9, Additional file 1). Using just the OMCL algorithm still obscured approximately 96 core functionalities by differential clustering; 73 of which could be traced back to probable splits in orthologous gene clusters of which half appear to be contributed to by the most evolutionary distant species, *P. propionicum*. Therefore, a functional-annotation of the pan-genomic matrix was generated to complement the orthologous matrix in a two-step process (S5, Additional file 4). We first utilised MATLAB to computationally combine entries with exact name matches. Secondly, the matrix was manually perused to correct entries which were similarly annotated but not exact matches due to a lack of annotation convention. This process was assisted by utilising EC numbers and primarily the MetaCyc database [71] to provide lists of synonymous enzymatic names. Little effort was spent curating hypothetical and putative proteins as these were expected to contribute negligibly to the functional understanding of *Propionibacterium* metabolism. Overall, the functional annotation collapsed the pan-genomic matrix by about two thirds while the number of core clusters remained roughly constant. Visualization of the orthologous and functional pan-genomic matrices was performed using Cluster 3.0 [72] and TreeView [73] (Figures 5 and 6, S1, Additional file 1).

Figure 5

Genome-scale metabolic model and Pan-GEM construction

The original annotations of the complete genomes were further processed through the KBase pipeline to draft-reconstructions (Figure 5). Due to the large number of *P. acnes* strains with little functional variability, suggested to be around 3–4% by the functional pan-genomic matrix, a model was only generated for *P. acnes* 6609. At the time of writing, it was not possible to either manually curate models using KBase or to upload a working, manually curated model into the framework which made it difficult to utilise any additional features of the pipeline. We instead utilised chemically defined media formulations from experiments or the literature to initially gap-fill reconstructions using the standard biomass template reaction; or complete media formulations where these were not available. A chemically defined media was developed in house for *P. acidipropionici* which included several trace elements, glucose as a carbon source, nitrate as a nitrogen source and phosphate as a source of phosphorous and the vitamins biotin, pantothenate, riboflavin, thiamin and vitamin B12.

A consensus biomass equation was next constructed by datamining the available literature. Most investigations have focussed on the unusual lipid and cell wall composition [74–76] of propionibacteria but other studies on the macromolecular composition of *Propionibacterium* were absent. We therefore composed a consensus biomass equation for *Propionibacterium* that captures the predominant composition of the genus based on current knowledge, while acknowledging that there will be significant

variability between different species; for example, the peptidoglycan structure is variable in the genus. DNA composition was determined from the genome sequence and RNA composition was estimated based on the *P. acidipropionici* 4875 RNA sequencing data. Where unknown, the macromolecular composition and protein compositions were complemented with quantitative data from *Streptomyces coelicolor* [77]. After finalising a biomass equation (S10, Additional file 6), previously added non-essential gap-fills were removed.

The pan-GEM represents the complete metabolic functionality of the individual reconstructions generated for each species. To create the pan-GEM, the individual metabolic reactions from each model are consolidated into a single supra-model with conserved associated gene annotations (S3, Additional file 3).

Pan-genome guided genome-scale metabolic reconstruction

The draft reconstructions developed were of low quality, containing numerous incorrect annotations and lacking key central carbon metabolic reactions including reactions of the Wood-Werkman cycle, highlighting the shortcomings of utilising popular automated annotation pipelines for non-model organisms. A combined pan-genomic guided annotation approach was used to functionally correct the annotations by incorporating functional information contained within the models, pan-genomic information and available data and is summarised in Figure 5. Briefly, the metabolic models gave functional cues to guide the annotation, such as gap-fills, blocked reactions, differing metabolic capabilities between organisms or testing catabolic capacity for experimentally verified nutrient sources including phenotype array information. Metabolism was explored on a pathway basis using pathways defined in MetaCyc [71] or subsystems from RAST to guide the search for potentially missing functionalities. Functional cues also come from the literature which contains strain specific information for certain genes and enzymatic capabilities. Another primary source of annotation cues comes from genomic-context of the orthologous and functional matrices. The genomic-context benefits the annotation by enhancing resolution and allows the identification of inconsistent annotations, potentially missed or incorrect annotations when they are present or absent in a small number of species and provides additional evidence for an annotation; such as if a function is flanked by genetic elements with related functions in some strains or highly conserved genomic regions. Analysis of the functional orthologous matrix allows the identification of reactions which were not captured in the model due to pipeline limitations and helps identify incorrect annotations such as inconsistent numbers of subunits of a particular complex. Potentially absent or incorrectly annotated genes were then probed by identifying sequences from closely related strains and by using BLAST to compare these sequences against a local database. This database contained all modelled *Propionibacterium* protein sequences and genomes and could be used to query the propionibacteria pan-proteome or look for unannotated sequences in the pan-genome. Results are then validated against orthologous gene clusters and functionality is assigned on a cluster basis.

This manual curation approach was applied finitely and focussed on pathways of the central carbon metabolism, particularly around the central phosphoenolpyruvate-pyruvate node and eventually expanded to include the respiratory metabolism and more specific metabolic traits of *Propionibacterium* including polyphosphate and trehalose metabolism. Particular attention was also given to amino acid metabolism which can influence fermentation products [78]. By referencing the literature and the BRENDA database [79], known enzyme promiscuities that could fill functional gaps were identified, cofactor usage of reactions in the model were re-assessed and reaction reversibility was adjusted based on biochemical evidence and modelling results. For example, the irreversibility of the pyrophosphate generating reactions through the actions of intracellular pyrophosphatases was set in all models. For a brief guide of the manual curation of genome-scale metabolic models from the RAST framework, refer to S11 (Additional file 1).

Phenotype array data generation and integration

Phenotypic growth array data (S6, Additional file 5) for *P. acidipropionici* 4875 was generated and used to refine the model using an in-house gap-filling algorithm (manuscript in preparation). The phenotype array and methodology was provided by Biolog, Inc. and was performed in accordance with their *Streptomyces* protocol with the modification of 2 µg/mL of known essential *Propionibacterium* vitamins; thiamine, biotin, pantothenic acid, riboflavin and vitamin B12. All additives to the array were defined except the proprietary inoculating fluids. *P. acidipropionici* 4875 grew on 110 of the 189 carbon sources tested of which 91 could be mapped to compounds in the ModelSEED database. Transporters for these 91 compounds were then added to the *P. acidipropionici* 4875 model. Next, the complete database of ModelSEED reactions was used to test if each of these compounds could sustain growth and where not possible, they were discarded from the analysis (14). Essential reactions required for growth on any of the conditions were then identified and added to the analysis. Finally, the minimal number of reactions needed to sustain growth across all conditions was simultaneously computed using a mixed-integer linear programming formulation. A final round of manual curation was performed such that all reactions incorporated into the model based on this work were either supported by genes, enzyme promiscuity information, were essential or otherwise the only reasonable candidate and results were propagated across other reconstructions where supported by genetic or biochemical evidence. The degradation pathways of phenylalanine and two related aromatic compounds completely lacked any genetic support for either of the currently known pathways for degradation and were therefore excluded.

Clustergram generation

Comparison of the biosynthetic and catabolic potential of amino acids in propionibacteria was performed by generating clustergrams. Reactions associated with the catabolism or anabolism of each amino acid were grouped and their presence or absence scored in each reconstruction. Differences between strains were quantified using the Jaccardian similarity coefficient. Details of these pathways, brief functional assessments and Jaccard score computation can be found in S12, Additional file 7. The

Jaccardian similarity coefficient was utilised as a distance metric to generate the clustergrams in MATLAB.

Bioreactor cultures and 'omics analysis

Propionibacterium acidipropionici ATCC 4875 was obtained from the American Type Culture Collection. Cells stored at -80°C , were used to inoculate serum bottles containing 50 mL PAM with 40 g/L of the relevant carbon source. Cultures were incubated statically at 32°C for 24 hours and used to inoculate Applikon Ez-Control bioreactors containing PAM and 100g/L of glucose, sucrose or glycerol. All experiments were performed in duplicates. Reactors were maintained at 32°C , pH 6.5 with 5 M NaOH, agitated at 200 rpm and the head space was degassed with nitrogen to maintain anaerobic conditions throughout the fermentation. Regular samples were taken to measure optical density using a Biochrom Libra S12 spectrophotometer and for organic acid analysis. Cells were harvested for RNA and protein extraction during mid-exponential phase. Total RNA and mRNA enrichment, sequence alignment and differential expression analysis was performed as described in [80]. For proteomics, cell pellets were lysed and trypsin digestion was performed as previously described in [81] and SWATH based quantification was performed as described in [82]. The differential abundance of proteins with more than 2 peptides with a 95% confidence score and a false discovery rate less than 1% was assessed using the LIMMA package. Extracellular organic acids, carbohydrates and alcohols were quantified by ion-exclusion chromatography as described in [27].

PFOR Knockout

The PFOR encoded in *nifJ1* gene (PFREUD_RS00925, CDS.199 this study) was knocked out in *P. freudenreichii* subsp. *shermanii* using CRISPR/Cas9 by interruption of the gene with gfpUV. Briefly, we assembled plasmid pPAC_Cas9 using pRG01 as the plasmid backbone and introducing the lacZ-MCS region, an erythromycin resistance gene (*ermE*) and a GC-optimised Cas9 under the control of promoter PermE. Gibson assembly was then used to create the knockout vector, pCas9_nifJ (Figure 6), by adding homologous arms designed to span 1 kb up and downstream of a PAM region identified in the *nifJ* gene, the gfpUV sequence and gRNA under the control of the P130 promoter from propionibacteria. The homologous arms and gfpUV were PCR amplified from *P. freudenreichii* subsp. *shermanii* and the pBRPprp_gfpuv plasmid respectively. Primer sequences and the 20-nt region for the gRNA are shown in Table 5. All cloning was performed in *E. coli* DH5 α (Bioline). PCR reactions were performed using Phusion Taq (NEB). Introduction of plasmid pCas9_nifJ into *P. freudenreichii* subsp. *shermanii* was performed by electroporation as previously described [28]. After 3 h recovery on PAM media (sucrose, 40 g/L; trypticase soy, 5 g/L; yeast extract, 10 g/L; K_2HPO_4 , 0.05 g/L; MnSO_4 , 0.05 g/L), cells were plated on NLB (sodium lactate, 10 g/L; trypticase soy, 10 g/L; yeast extract, 10 g/L; glucose, 10 g/L) agar and incubated in an anaerobic jar at 30 C. After 10 days, colonies were identified and tested by colony PCR to confirm the successful knockout. Positive Δ nifJ colonies were then grown on NLB liquid media for 72 h.

Figure 6

List Of Abbreviations

GEM: genome-scale metabolic model

GPR: gene-protein-reaction associations

PFOR: pyruvate flavodoxin oxidoreductase

P. acidipropionici: *Propionibacterium acidipropionici*

PAM media: media used in this study as defined in the Methods section

DW HA: downstream homologous arms

UP HA: upstream homologous arms

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

All data is available in the Additional files.

Competing interests

Dow has commercial interest in commercialising propionic acid production. CS is an employee of Corteva, formerly Dow Agrosiences.

Author's contributions

TM performed the bioinformatics analyses, generated and curated the genome-scale metabolic reconstructions and wrote the manuscript. RAGG performed the knockout study in *P. freudenreichii* subsp. *shermanii*. RWP provided support for bioinformatic analyses including analysis of transcriptomic data.

CS performed the BIOLOG analysis. LKN and EM provided intellectual guidance and contributed to the manuscript. EM performed the fermentations of *P. acidipropionici* on different carbon sources and prepared samples for 'omics analysis. All authors contributed in writing of the manuscript and read and approved the final manuscript.

Funding

The research was funded by the Australian Research Council *via* a linkage grant (LP120100517) with Dow as the commercial partner. The funding agency had no involvement in the design of the experiments nor was involved in analysing the data. Dow has interest in commercialising propionic acid production. TM was supported by an APA scholarship. EM was funded by the Queensland government by an Accelerate Fellowship.

Acknowledgements

We thank Paul Swanson for his assistance performing and providing the BIOLOG analysis. This research was supported by the Queensland Node of Metabolomics Australia at The University of Queensland, a Bioplatforms Australia funded initiative under NCRIS.

References

1. Seeliger S, Janssen PH, Schink B: Energetics and kinetics of lactate fermentation to acetate and propionate via methylmalonyl-CoA or acrylyl-CoA. FEMS Microbiol Lett 2002, 211:65–70.
2. DE VRIES W, van WYCK-KAPTEYN WMC, STOUTHAMER AH: Generation of ATP during Cytochrome-linked Anaerobic Electron Transport in Propionic Acid Bacteria. J Gen Microbiol 1973, 76:31–41.
3. Rodriguez B a., Stowers CC, Pham V, Cox BM: The production of propionic acid, propanol and propylene via sugar fermentation: an industrial perspective on the progress, technical challenges and future outlook. Green Chem 2014, 16:1066.
4. Gonzalez-Garcia R, McCubbin T, Navone L, Stowers C, Nielsen L, Marcellin E: Microbial Propionic Acid Production. Fermentation 2017, 3:21.
5. Corvec S: Clinical and Biological Features of Cutibacterium (Formerly Propionibacterium) avidum, an Underrecognized Microorganism. Clin Microbiol Rev 2018, 31.
6. Chavali AK, D'Auria KM, Hewlett EL, Pearson RD, Papin JA: A metabolic network approach for the identification and prioritization of antimicrobial drug targets. Trends Microbiol 2012, 20:113–123.
7. Deutsch S-M, Le Bivic P, Hervé C, Madec M-N, LaPointe G, Jan G, Le Loir Y, Falentin H: Correlation of the capsular phenotype in Propionibacterium freudenreichii with the level of expression of gtf, a unique polysaccharide synthase-encoding gene. Appl Environ Microbiol 2010, 76:2740–6.
8. Zhuge X, Liu L, Shin H, Chen RR, Li J, Du G, Chen J: Development of a Propionibacterium-Escherichia coli shuttle vector for metabolic engineering of Propionibacterium jensenii, an efficient producer of

- propionic acid. *Appl Environ Microbiol* 2013, 79:4595–602.
9. Jiang L, Cui H, Zhu L, Hu Y, Xu X, Li S, Huang H: Enhanced propionic acid production from whey lactose with immobilized *Propionibacterium acidipropionici* and the role of trehalose synthesis in acid tolerance. *Green Chem* 2015, 17:250–259.
 10. Dohm JC, Lottaz C, Borodina T, Himmelbauer H: Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res* 2008, 36:e105–e105.
 11. Angelova M, Kalajdziski S, Kocarev L: Computational Methods for Gene Finding in Prokaryotes. In *Web Proceedings, ISSN 2010; 2010:11–20*.
 12. Suwannakham S, Huang Y, Yang S-T: Construction and characterization of *ack* knock-out mutants of *Propionibacterium acidipropionici* for enhanced propionic acid fermentation. *Biotechnol Bioeng* 2006, 94:383–95.
 13. Falentin H, Deutsch S-M, Jan G, Loux V, Thierry A, Parayre S, Maillard M-B, Dherbécourt J, Cousin FJ, Jardin J, Siguier P, Couloux A, Barbe V, Vacherie B, Wincker P, Gibrat J-F, Gaillardin C, Lortal S: The complete genome of *Propionibacterium freudenreichii* CIRM-BIA1, a hardy actinobacterium with food and probiotic applications. *PLoS One* 2010, 5:e11748.
 14. Parizzi LP, Grassi MCB, Llerena L a, Carazzolle MF, Queiroz VL, Lunardi I, Zeidler AF, Teixeira PJPL, Mieczkowski P, Rincones J, Pereira GAG: The genome sequence of *Propionibacterium acidipropionici* provides insights into its biotechnological and industrial potential. *BMC Genomics* 2012, 13:562.
 15. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R: The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* 2014, 42:206–214.
 16. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O: The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 2008, 9:75.
 17. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR, Xia F: RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 2015, 5:8365.
 18. Magnúsdóttir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, Noronha A, Greenhalgh K, Jäger C, Baginska J, Wilmes P, Fleming RMT, Thiele I: Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat Biotechnol* 2016, 35:81–89.
 19. Gao X-Y, Zhi X-Y, Li H-W, Klenk H-P, Li W-J: Comparative Genomics of the Bacterial Genus *Streptococcus* Illuminates Evolutionary Implications of Species Groups. *PLoS One* 2014, 9:e101229.
 20. Wassenaar TM, Jun S-R, Wanchai V, Patumcharoenpol P, Nookaew I, Schlum K, Leuze MR, Ussery DW: Insights from Comparative Genomics of the Genus *Salmonella*. In *Current Topics in Salmonella and Salmonellosis*. InTech; 2017.

21. Steinway SN, Biggs MB, Loughran TP, Papin JA, Albert R: Inference of Network Dynamics and Metabolic Interactions in the Gut Microbiome. *PLOS Comput Biol* 2015, 11:e1004338.
22. Driscoll TP, Verhoeve VI, Guillotte ML, Lehman SS, Rennoll SA, Beier-Sexton M, Rahman MS, Azad AF, Gillespie JJ: Wholly Rickettsia! Reconstructed Metabolic Profile of the Quintessential Bacterial Parasite of Eukaryotic Cells. *MBio* 2017, 8:e00859–17.
23. Willenbrock H, Hallin PF, Wassenaar TM, Ussery DW: Characterization of probiotic *Escherichia coli* isolates with a novel pan-genome microarray. *Genome Biol* 2007, 8:R267.
24. Wang Z, Ammar EM, Zhang A, Wang L, Lin M, Yang S-T: Engineering *Propionibacterium freudenreichii* subsp. *shermanii* for enhanced propionic acid fermentation: Effects of overexpressing propionyl-CoA:Succinate CoA transferase. *Metab Eng* 2015, 27:46–56.
25. Liu L, Zhuge X, Shin H, Chen RR, Li J, Du G, Chen J: Improved Production of Propionic Acid in *Propionibacterium jensenii* via Combinational Overexpression of Glycerol Dehydrogenase and Malate Dehydrogenase from *Klebsiella pneumoniae*. *Appl Environ Microbiol* 2015, 81:2256–2264.
26. Luna-Flores, C. H.; Stowers, C. C.; Cox, B.; Nielsen, L. K.; Marcellin E: Linking genotype and phenotype in an economically viable propionic acid biosynthesis process. *Biotechnol Biofuels* 2018, 11, 224.
27. Luna-Flores CH, Palfreyman RW, Krömer JO, Nielsen LK, Marcellin E: Improved production of propionic acid using genome shuffling. *Biotechnol J* 2017, 12:1600120.
28. Navone L, McCubbin T, Gonzalez-Garcia RA, Nielsen LK, Marcellin E: Genome-scale model guided design of *Propionibacterium* for enhanced propionic acid production. *Metab Eng Commun* 2018, 6:1–12.
29. Stowers CC, Cox BM, Rodriguez B a: Development of an industrializable fermentation process for propionic acid production. *J Ind Microbiol Biotechnol* 2014, 41:837–852.
30. Bott M, Pfister K, Burda P, Kalbermatter O, Woehlke G, Dimroth P: Methylmalonyl-CoA Decarboxylase from *Propionigenium Modestum*. Cloning and Sequencing of the Structural Genes and Purification of the Enzyme Complex. *Eur J Biochem* 1997, 250:590–599.
31. Feng X, Chen F, Xu H, Wu B, Li H, Li S, Ouyang P: Green and economical production of propionic acid by *Propionibacterium freudenreichii* CCTCC M207015 in plant fibrous-bed bioreactor. *Bioresour Technol* 2011, 102:6141–6.
32. Matsakas L, Hrůzová K, Rova U, Christakopoulos P: Biological Production of 3-Hydroxypropionic Acid: An Update on the Current Status. *Fermentation* 2018, 4:13.
33. Toraya T, Kuno S, Fukui S: Distribution of coenzyme B12-dependent diol dehydratase and glycerol dehydratase in selected genera of Enterobacteriaceae and Propionibacteriaceae. *J Bacteriol* 1980, 141:1439–42.
34. Falentin H, Deutsch S-M, Jan G, Loux V, Thierry A, Parayre S, Maillard M-B, Dherbécourt J, Cousin FJ, Jardin J, Siguier P, Couloux A, Barbe V, Vacherie B, Wincker P, Gibrat J-F, Gaillardin C, Lortal S: The Complete Genome of *Propionibacterium freudenreichii* CIRM-BIA1T, a Hardy Actinobacterium with Food and Probiotic Applications. *PLoS One* 2010, 5:e11748.

35. Furuichi K, Hojo K, Katakura Y, Ninomiya K, Shioya S: Aerobic culture of *Propionibacterium freudenreichii* ET-3 can increase production ratio of 1,4-dihydroxy-2-naphthoic acid to menaquinone. *J Biosci Bioeng* 2006, 101:464–70.
36. Cheng S, Fan C, Sinha S, Bobik TA: The PduQ Enzyme Is an Alcohol Dehydrogenase Used to Recycle NAD⁺ Internally within the Pdu Microcompartment of *Salmonella enterica*. *PLoS One* 2012, 7:e47144.
37. Lee INNHEE, Fredrickson AG, Tsuchiya HM: Diauxic Growth of *Propionibacterium shermanii*. 1974, 28:831–835.
38. Bott M, Niebisch A: The respiratory chain of *Corynebacterium glutamicum*. *J Biotechnol* 2003, 104:129–153.
39. Chai Y, Kolter R, Losick R: A Widely Conserved Gene Cluster Required for Lactate Utilization in *Bacillus subtilis* and Its Involvement in Biofilm Formation. *J Bacteriol* 2009, 191:2423–2430.
40. Pinchuk GE, Rodionov D a, Yang C, Li X, Osterman AL, Dervyn E, Geydebrekht O V, Reed SB, Romine MF, Collart FR, Scott JH, Fredrickson JK, Beliaev AS: Genomic reconstruction of *Shewanella oneidensis* MR-1 metabolism reveals a previously uncharacterized machinery for lactate utilization. *Proc Natl Acad Sci U S A* 2009, 106:2874–9.
41. Sone N: The redox reactions in propionic acid fermentation. 3. Enzymatic properties of NAD-independent glycerol-phosphate dehydrogenase from *Propionibacterium arabinosum*. *J Biochem* 1973, 74:297–305.
42. Schwartz AC, Sporkenbach J: The electron transport system of the anaerobic *Propionibacterium shermanii*. *Arch Microbiol* 1975, 102:261–273.
43. Rodriguez B a., Stowers CC, Pham V, Cox BM: The production of propionic acid, propanol and propylene via sugar fermentation: an industrial perspective on the progress, technical challenges and future outlook. *Green Chem* 2014, 16:1066–1076.
44. Crow VL: Utilization of Lactate Isomers by *Propionibacterium freudenreichii* subsp. *shermanii*: Regulatory Role for Intracellular Pyruvate. *Appl Environ Microbiol* 1986, 52:352–8.
45. Elbehti A, Brasseur G, Lemesle-Meunier D: First evidence for existence of an uphill electron transfer through the bc(1) and NADH-Q oxidoreductase complexes of the acidophilic obligate chemolithotrophic ferrous ion-oxidizing bacterium *Thiobacillus ferrooxidans*. *J Bacteriol* 2000, 182:3602–6.
46. Lückner S, Nowka B, Rattei T, Spieck E, Daims H: The Genome of *Nitrospina gracilis* Illuminates the Metabolism and Evolution of the Major Marine Nitrite Oxidizer. *Front Microbiol* 2013, 4(February):27.
47. Weghoff MC, Bertsch J, Müller V: A novel mode of lactate metabolism in strictly anaerobic bacteria. *Environ Microbiol* 2015, 17:670–7.
48. Ledbetter RN, Garcia Costas AM, Lubner CE, Mulder DW, Tokmina-Lukaszewska M, Artz JH, Patterson A, Magnuson TS, Jay ZJ, Duan HD, Miller J, Plunkett MH, Hoben JP, Barney BM, Carlson RP, Miller A-F, Bothner B, King PW, Peters JW, Seefeldt LC: The Electron Bifurcating FixABCX Protein Complex from

- Azotobacter vinelandii*: Generation of Low-Potential Reducing Equivalents for Nitrogenase Catalysis. *Biochemistry* 2017, 56:4177–4190.
49. Buckel W, Thauer RK: Energy conservation via electron bifurcating ferredoxin reduction and proton/Na(+) translocating ferredoxin oxidation. *Biochim Biophys Acta* 2013, 1827:94–113.
 50. Chiba Y, Kamikawa R, Nakada-Tsukui K, Saito-Nakano Y, Nozaki T: Discovery of PPi-type Phosphoenolpyruvate Carboxykinase Genes in Eukaryotes and Bacteria. *J Biol Chem* 2015, 290:23960–70.
 51. Brzuszkiewicz E, Weiner J, Wollherr A, Thürmer A, Hüpeden J, Lomholt HB, Kilian M, Gottschalk G, Daniel R, Mollenkopf H-J, Meyer TF, Brüggemann H: Comparative genomics and transcriptomics of *Propionibacterium acnes*. *PLoS One* 2011, 6:e21581.
 52. Zhang H, Ishige K, Kornberg A: A polyphosphate kinase (PPK2) widely conserved in bacteria. *Proc Natl Acad Sci U S A* 2002, 99:16678–83.
 53. Zhang Q, Li Y, Tang CM: The role of the exopolyphosphatase PPX in avoidance by *Neisseria meningitidis* of complement-mediated killing. *J Biol Chem* 2010, 285:34259–68.
 54. Cardoso FS, Castro RF, Borges N, Santos H: Biochemical and genetic characterization of the pathways for trehalose metabolism in *Propionibacterium freudenreichii*, and their role in stress response. *Microbiology* 2007, 153:270–280.
 55. Thierry A, Maillard M-B, Yvon M: Conversion of L-leucine to isovaleric acid by *Propionibacterium freudenreichii* TL 34 and ITGP23. *Appl Environ Microbiol* 2002, 68:608–15.
 56. Carlier JP, Sellier N: Gas chromatographic-mass spectral studies after methylation of metabolites produced by some anaerobic bacteria in spent media. *J Chromatogr* 1989, 493:257–73.
 57. Cotter PD, Hill C: Surviving the Acid Test: Responses of Gram-Positive Bacteria to Low pH. *Microbiol Mol Biol Rev* 2003, 67:429–453.
 58. Contreras-Moreira B, Vinuesa P: GET_HOMOLOGUES, a Versatile Software Package for Scalable and Robust Microbial Pangenome Analysis. *Appl Environ Microbiol* 2013, 79:7696–7701.
 59. Arkin AP, Stevens RL, Cottingham RW, Maslov S, Henry CS, Dehal P, Ware D, Perez F, Harris NL, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Gunter D, Murphy-Olson D, Chan S, Kamimura RT, Brettin TS, Meyer F, Chivian D, Weston DJ, Glass EM, Davison BH, Kumari S, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, et al.: The DOE Systems Biology Knowledgebase (KBase). *bioRxiv* 2016.
 60. Luna-Flores CH, Nielsen LK, Marcellin E: Genome Sequence of *Propionibacterium acidipropionici* ATCC 55737. *Genome Announc* 2016, 4:e00248–16.
 61. Ordogh L, Hunyadkurti J, Voros A, Horvath B, Szucs A, Urban E, Kereszt A, Kondorosi E, Nagy I: Complete Genome Sequence of *Propionibacterium avidum* Strain 44067, Isolated from a Human Skin Abscess. *Genome Announc* 2013, 1:e00337–13-e00337–13.
 62. Human Microbiome Project Consortium: Structure, function and diversity of the healthy human microbiome. *Nature* 2012, 486:207–14.

63. Human Microbiome Project Consortium: A framework for human microbiome research. *Nature* 2012, 486:215–21.
64. Brüggemann H, Henne A, Hoster F, Liesegang H, Wiezer A, Strittmatter A, Hujer S, Dürre P, Gottschalk G: The complete genome sequence of *Propionibacterium acnes*, a commensal of human skin. *Science* (80-) 2004, 305:671–3.
65. Hunyadkürti J, Feltóti Z, Horváth B, Nagymihály M, Vörös A, McDowell A, Patrick S, Urbán E, Nagy I: Complete genome sequence of *Propionibacterium acnes* type IB strain 6609. *J Bacteriol* 2011, 193:4561–2.
66. Vörös A, Horváth B, Hunyadkürti J, McDowell A, Barnard E, Patrick S, Nagy I: Complete genome sequences of three *Propionibacterium acnes* isolates from the type IA(2) cluster. *J Bacteriol* 2012, 194:1621–2.
67. Horváth B, Hunyadkürti J, Vörös A, Fekete C, Urbán E, Kemény L, Nagy I: Genome sequence of *Propionibacterium acnes* type II strain ATCC 11828. *J Bacteriol* 2012, 194:202–3.
68. Minegishi K, Aikawa C, Furukawa A, Watanabe T, Nakano T, Ogura Y, Ohtsubo Y, Kurokawa K, Hayashi T, Maruyama F, Nakagawa I, Eishi Y: Complete Genome Sequence of a *Propionibacterium acnes* Isolate from a Sarcoidosis Patient. *Genome Announc* 2013, 1.
69. Fitz-Gibbon S, Tomida S, Chiu B-H, Nguyen L, Du C, Liu M, Elashoff D, Erfe MC, Loncaric A, Kim J, Modlin RL, Miller JF, Sodergren E, Craft N, Weinstock GM, Li H: *Propionibacterium acnes* strain populations in the human skin microbiome associated with acne. *J Invest Dermatol* 2013, 133:2152–60.
70. Li L, Stoeckert CJ, Roos DS: OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 2003, 13:2178–89.
71. Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, Holland TA, Keseler IM, Kothari A, Kubo A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Subhraveti P, Weaver DS, Weerasinghe D, Zhang P, Karp PD: The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* 2014, 42(Database issue):D459–71.
72. Hoon MJL De, Imoto S, Nolan J, Miyano S: Open source clustering software. 2004, 20:1453–1454.
73. Page RD: TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 1996, 12:357–8.
74. Sutcliffe IC, Shaw N: An inositol containing lipomannan from *Propionibacterium freudenreichii*. *FEMS Microbiol Lett* 1989, 59:249–251.
75. Cummins CS, White RH: Isolation, identification, and synthesis of 2,3-diamino–2,3-dideoxyglucuronic acid: a component of *Propionibacterium acnes* cell wall polysaccharide. *J Bacteriol* 1983, 153:1388–93.
76. Moss CW, Dowell VR, Farshchi D, Raines LJ, Cherry WB: Cultural characteristics and fatty acid composition of propionibacteria. *J Bacteriol* 1969, 97:561–70.

77. Kim M, Sang Yi J, Kim J, Kim J-N, Kim MW, Kim B-G: Reconstruction of a high-quality metabolic model enables the identification of gene overexpression targets for enhanced antibiotic production in *Streptomyces coelicolor* A3(2). *Biotechnol J* 2014, 9:1185–94.
78. Piveteau P: Metabolism of lactate and sugars by dairy propionibacteria: A review. *Lait* 1999, 79:23–41.
79. Scheer M, Grote A, Chang A, Schomburg I, Munaretto C, Rother M, Söhngen C, Stelzer M, Thiele J, Schomburg D: BRENDA, the enzyme information system in 2011. *Nucleic Acids Res* 2011, 39(Database issue):D670–6.
80. Marcellin E, Mercer TR, Licona-Cassani C, Palfreyman RW, Dinger ME, Steen JA, Mattick JS, Nielsen LK: *Saccharopolyspora erythraea*'s genome is organised in high-order transcriptional regions mediated by targeted degradation at the metabolic switch. *BMC Genomics* 2013, 14:15.
81. Licona-Cassani C, Lim S, Marcellin E, Nielsen LK: Temporal Dynamics of the *Saccharopolyspora erythraea* Phosphoproteome. *Mol Cell Proteomics* 2014, 13:1219–1230.
82. Chang RYK, Etheridge N, Nouwens AS, Dodd PR: SWATH analysis of the synaptic proteome in Alzheimer's disease. *Neurochem Int* 2015, 87:1–12.
83. Kiatpapan P, Hashimoto Y, Nakamura H, Piao YZ, Ono H, Yamashita M, Murooka Y: Characterization of pRGO1, a plasmid from *Propionibacterium acidipropionici*, and its use for development of a host-vector system in propionibacteria. *Appl Environ Microbiol* 2000, 66:4688–95.

Tables

Due to technical limitations, all tables are only available for download from the Supplementary Files section.

Supplementary Files Legend

Additional file 1: S1, S4, S7-9, S11 (.docx)

S1. Additional 6 figures complementing the pan-genomic analysis including estimates and representations of the core and pan-genome size, phylogenetic relationship and heat maps of the orthologous and functional clusters. **S4.** Visualization of transcriptomics data associated with gene clusters of interest; the putative ferredoxin bifurcating enzymes associated with the pyruvate-ferredoxin oxidoreductase, the sodium-ion pumping methylmalonyl-CoA decarboxylase and the pyruvate oxidase. **S7.** Analysis of the influence of contigs on the pan-genomic and pan-GEM approaches. **S8.** Analysis of the quality of gene-calling algorithms and influence of the choice of the algorithm on the initial reconstruction. **S9.** Representations of the robust core and pan-genomes and illustration of the loss of information of functional capabilities when this type of analysis is used. **S11.** A brief guide detailing important steps and useful strategies to assist in the manual curation of genome-scale metabolic reconstructions resulting from automated pipelines, particularly RAST/KBase.

Additional file 2: S2 (.xlsx)

S2. Phenotype data collected from *P. acidipropionici* 4875 grown on three different carbon sources, including extracellular metabolite concentrations, proteomics and metabolomics data.

Additional file 3: S3 (.xlsx, .xml, .mat and .gbk)

S3. Final Pan-GEM model and original GEMs resulting from the reconstruction pipeline, corresponding .xml and .mat files and associated GenBank files in a zipped folder.

Additional file 4: S5 (.xlsx)

S5. Pan-genomic matrix representations based on orthologous clusters and partial and full functional compactions; used to guide manual curation of the model.

Additional file 5: S6 (.xlsx)

S6. BIOLOG data for *P. acidipropionici* 4875 used to guide gap-filling work and manual curation.

Additional file 6: S10 (.xlsx)

S10. Consensus biomass equation reconstruction.

Additional file 7: S12 (.xlsx)

S12. Definition and analysis for presence/absence of the amino acid catabolic and anabolic pathways in various species and computation of the Jaccard distances.

Figures

A - LACTATE METABOLISM IN *PROPIONIBACTERIUM*

B - OPTIONS FOR OVERCOMING THE REDOX ENERGY BARRIER

1 REVERSE ELECTRON TRANSPORT

2 ELECTRON BIFURCATING LACTATE DEHYDROGENASE

3 PUTATIVE ELECTRON BIFURCATING COMPLEX PROPOSED

Figure 1

A novel ferredoxin-based energy conservation mechanism explains discrepancies between the literature and observed lactate metabolism in *Propionibacterium*. A- catabolic pathways for the conversion of lactate to propionate and acetate are illustrated and are most representative of the dairy propionibacteria, particularly *P. acidipropionici*, where the 2:1 ratio of propionate to acetate has been observed. The anaerobic electron transport chain is illustrated (bordered in red) and can be entered at complex II

(cytochrome B) via quinols, presuming translocation of 2 protons or half an ATP gain in energy, or complex I via NADH netting an additional 2 protons pumped. The balance over lactate catabolism illustrates an overall redox imbalance exists where a quinol must be used to regenerate an NADH. B-potential strategies for overcoming this redox challenge are illustrated and are referenced to the appropriate location in metabolism in A. There was no genetic evidence for the first two mechanisms in propionibacteria, however the last step is proposed based on similarities with flanking genes of the pyruvate ferredoxin oxidoreductase and the Nfn complex of *Clostridium kluyveri* which could allow for the regeneration of NADH from a quinone.

Figure 2

Acetate, acetoin and ethanol metabolism of *Propionibacterium*. (Black) arrows represent reactions present in at least one species. (Red) arrows represent reaction in all species (core).

Figure 3

Diagrams of the pan-genomic nitrate (A), sulphate (B) and trehalose (C) metabolism in *Propionibacterium*. (Black) arrows represent reactions present in at least one species. (Red) arrows represent reaction in all species (core). *Enzyme is present in both *P. acidipropionici* strains but 4875 has a frameshift mutation likely rendering the enzyme non-functional.

Figure 4

Heat map and clustering of amino acid biosynthesis (A) and degradation (B) pathways using the Jaccard similarity metric. In this metric, 0 represents no reactions in common and 1 represents all reactions shared. Plots to the left are referenced to the pan-reactome (distances represent the proportion of all reactions of the pan-genome ascribed to that pathway present in the organism) while plots to the right represent the same information referenced to *P. acidipropionici* 4875. Commensal strains are observed to

have a reduced set of biosynthetic capabilities for amino acids and cluster out from the dairy strains. Extensive variability is observed in the arginine, methionine and proline degradation pathways. The reductive arginine degradation pathway is observed in *P. acidipropionici* while the oxidative pathway is observed in commensal species, possibly reflecting adaptation to aerobic conditions. Histidine and tryptophan degradation capabilities are observed only in selected commensal strains.

Figure 5

Graphical depiction of the bioinformatics workflow. Improved annotations of the genome, captured as a genome-scale metabolic reconstruction, are pursued using both functional approaches and genomic-context approaches which capitalise on information contained within the pan-genome.

Figure 6

Plasmid map for pCas9_nifJ. Knockout is driven by recombination between downstream and upstream homologous arms (DW HA and UP HA respectively) and the nifJ1 gene in the chromosome, leading to disruption of the nifJ1 with gfpUV.

Supplementary Files

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

- PAC55737model.xml
- Additionalfile4S5.xlsx
- PropionibacteriumpropionicumF0230a.gbk
- Tables.docx
- Propionibacteriumacnes6609.gbk
- Propionibacteriumavidum44067.gbk
- Propionibacteriumfreudenreichiisubsp.shermaniiCIRMBIA1.gbk
- Propionibacteriumacidipropionici55737.gbk
- Additionalfile7S12.xlsx
- Additionalfile6S10.xlsx
- Additionalfile5S6.xlsx
- Additionalfile2S2.xlsx
- Pshermmodel.xml
- Additionalfile3S3.xlsx
- Pacnesmodel.mat
- Pshermmodel.mat
- Pacnesmodel.xml
- Pavidummodel.mat
- Additionalfile1S1S4S79S11.docx
- PAC4875model.xml
- PAC4875model.mat
- PAC55737model.mat
- Pavidummodel.xml
- Ppropionicummodel.xml
- PropionibacteriumacidipropioniciATCC4875.gbk
- Ppropionicummodel.mat