Identification of widely conserved biosynthetic gene cluster involved in pigment production of Bacillus subtilis

Stannius, R.O.; Dunlap, C.A.; Morvan, E.; Berbon, M.; Lecomte, S.; Loquet, A.; ... ; Yang Yu-Liang

mSystems | Open Peer Review | Bacteriology | Research Article

Identification of widely conserved biosynthetic gene cluster involved in pigment production of *Bacillus subtilis*

Rune Overlund Stannius,[1,2] Christopher A. Dunlap,[3] Estelle Morvan,[4] Mélanie Berbon,[5] Sophie Lecomte,[5] Antoine Loquet,[5] Ákos T. Kovács[1,2]

AUTHOR AFFILIATIONS See affiliation list on p. 12.

ABSTRACT *Bacillus subtilis* is widely studied in the microbial secondary metabolite (SM) field due to its rich variety of important natural products and genetic tractability. We report a pigment observed in *B. subtilis* soil isolate MB9_B4 on certain media. We characterize the conditions where this pigment is produced and identify the corresponding biosynthetic gene cluster (BGC) using a comparative genomic approach exploiting our strain collection containing other isolates with pigment production ability. The responsible BGC carried several genes, which were annotated as parts of the tryptophan biosynthesis pathway, possibly originating from duplication and divergence of originally primary metabolism. Identification of the pigment gene cluster additionally led to the discovery of additional pigment BGC carrier *B. subtilis* isolates, some of which were described at the earliest in 1896 under the name *Bacillus aterrimus*, referring to a characteristic dark pigmentation (the Latin "aterrimus" meaning very black). In addition, we employed solid-state nuclear magnetic resonance and Fourier transform infrared spectroscopies to characterize the chemical groups of the pigment. This study describes the chemical and biological features of a new class of SM BGC, which we hope will serve to improve the current BGC discovery pipelines in *Bacilli*.

IMPORTANCE Identification of novel microbial secondary metabolites (SMs) and their biosynthetic gene cluster (BGC) has become increasingly difficult, especially in *Bacilli*, as the tools for screening and genome mining are dependent on clear function or similarity to already known BGCs. Pigments are SMs identified by their absorption of visible light, resulting in a certain color perceived by our eyes at sufficient concentrations. Thereby, pigments provide evidence of a BGC without knowing the sequence or function. Expanding the known repertoire of SM BGCs with novel BGCs will further reinforce the identification of a broader set of BGCs by mining tools such as antiSMASH.

KEYWORDS *Bacillus subtilis*, pigment, melanin, biosynthetic gene cluster, genome-wide association study

Bacillus subtilis is a gram-positive soil-dwelling bacterium that, due to its genetic tractability and spore development, has been one of the focal species in microbiology with decades of research into its lifecycle and sporulation (1–4). Concomitantly with sporulation, many *Bacilli* produce a wide array of pigments with varying structures and proposed roles, mainly as UV protection and reactive oxygen species (ROS) scavenging benefitting spore resilience (5–9). However, pigments may have other roles beyond just their spectral properties and ability to scavenge ROS. In *B. subtilis*, the most well-studied pigment, pulcherriminic acid, has been shown to modulate iron availability by sequestering Fe(III) from competing organisms and preventing harmful Fenton reactions (10) and also influence biofilm development (11, 12). Pigments pose a promising target for study due to their intrinsic visibility without the need for specific detection technology.

Editor Yu-Liang Yang, Academia Sinica Agricultural Biotechnology Research Center, Tainan City, Taiwan

Address correspondence to Ákos T. Kovács, a.t.kovacs@biology.leidenuniv.nl.

The authors declare no conflict of interest.

See the funding table on p. 12.

Received 28 May 2025
Accepted 8 June 2025
Published 3 July 2025

Pigments belong to the group of secondary metabolites (SMs) that are differentiated from primary metabolism by having specialized functions not directly related to the growth and reproduction of the producing organism. SMs are a rich source of novel therapeutics and functional molecules for society, with a large majority of medicinal compounds either being SMs or derivatives of SMs (13, 14). However, the discovery of truly novel SMs is becoming difficult, especially in the *Bacillus* genus; screening assays often remove SMs from their context and test for specific functions, which further skews our knowledge toward well-known SM classes and functions while losing cryptically expressed SM biosynthetic gene clusters (BGCs) and restricting efforts to cultivable bacteria (15). One way to avoid reliance on expression and functional assays would be through sequence-based approaches, which can identify potential BGCs by similarity to already known clusters. Such an approach has the added advantage of being culture independent, allowing access to the largely untapped diversity of microorganisms in nature (13). However, sequence-based approaches are only likely to find SM BGCs that are similar in sequence to already defined BGC classes and therefore are unlikely to identify truly novel SM BGCs (16).

Evolutionarily, SM biosynthetic genes are thought to originate from primary metabolism through duplication and diversification of enzymes in which the encoded proteins gain new substrate specificity or enzymatic action (17). More recently diverged BGCs are still highly similar to their progenitor genes from the primary metabolism and are often not identified by various BGC-focused genome-mining tools as parts of the secondary metabolism. Thus, exploiting the evolutionary principles involved in the origin of BGCs has been proposed as a strategy with the genome-mining approach EvoMining, which circumvents aforementioned challenges by instead targeting enzyme family duplications and identifying potential BGCs by conserved genomic vicinity and phylogenetic reconstruction of genes encoding the duplicated enzymes (16, 18). Such an evolutionary approach is dependent on the likely assumption that SM biosynthetic genes are duplicated from the primary metabolism. However, the path from primary metabolism to SM is uncertain and difficult to observe in the laboratory due to the evolutionary timescale of such adaptation. Thus, both traditional similarity- and evolution-based approaches can benefit from additional examples of BGCs that are less diverged from the primary metabolism to improve the prediction ability of these novel methods and expand the catalogue for similarity searches.

In this study, we identify a dark water-soluble pigment produced by a *B. subtilis* soil isolate, MB9_B4. Through a genome-wide association study (GWAS), we were able to identify the putative gene cluster associated with pigment production, which consisted of multiple genes annotated as homologs of the primary metabolism for, among others, tryptophan. Blast searches revealed several other isolates that carry the pigment BGC, some of which had been studied for their pigmentation in different media conditions under the name *Bacillus aterrimus* as early as in 1896 (19–21). The chemical nature of the pigment was assessed by high-resolution solid-state nuclear magnetic resonance (NMR) and Attenuated total reflection Fourier transform infrared (ATR-FTIR) spectroscopy. This study adds to the knowledge base of less well-studied SMs in *Bacilli* and provides an example of a primary metabolism-derived SM, which could be useful for improving our BGC discovery pipelines.

MATERIALS AND METHODS

Strains, chemicals, and genetic modification

The list of *B. subtilis* strains used for the experiments and the additional isolates included in phylogeny can be found in Table S1. Lysogeny broth (LB; Lennox, Carl Roth, Karlsruhe, Germany; 10 g/L tryptone, 5 g/L yeast extract, and 5 g/L NaCl) was used for routine culturing. LBg (LB supplemented with 1% v/v glycerol), LBm (LB supplemented with 0.1 mmol/L $MnCl_2$), LBgm (LB supplemented with both 1% v/v glycerol and 0.1 mmol/L $MnCl_2$, based on reference 22), MSgg (1.925 mmol/L KH_2PO_4, 3.075 mmol/L K_2HPO_4,

adjusted to pH 7, 100 mmol/L 3-(N-morpholino)propanesulfonic acid [MOPS], 2 mmol/L $MgCl_2$, 700 µmol/L $CaCl_2$, 50 µmol/L $MnCl_2$, 50 µmol/L $FeCl_3$, 1 µmol/L $ZnCl_2$, 2 µmol/L thiamine, 0.5% v/v glycerol, 0.5% w/v L-glutamic acid potassium salt, 50 µg/ml tryptophan, [23]), Potato dextrose agar (PDA, Carl Roth, Karlsruhe, Germany), and Kings B agar (KB, Sigma Aldrich, Taufkirchen, Germany) were additionally used for assaying pigment production under various conditions. Antibiotics were used at the following final concentrations: spectinomycin (spec) 100 µg/mL, kanamycin (kan) 5 µg/mL, tetracycline (tet) 10 µg/mL, chloramphenicol (chl) 10 µg/mL, erythromycin (erm) 5 µg/mL for erm^R strains and 1 µg/mL combined with 12.5 µg/mL lincomycin for MLS^R strains (antibiotics were obtained from Carl Roth, Karlsruhe, Germany).

All genetic engineering was performed as described below using a modified version of the protocol described in reference 24. Briefly, overnight cultures were pelleted and resuspended in 100 µL MQ water. 10 µL of the concentrated cells was transferred into 2 mL competence medium (80 mmol/L K_2HPO_4, 38.2 mmol/L KH_2PO_4, 20 g/L glucose, 3 mmol/L Na_3-citrate, 45 µmol/L ferric NH_4-citrate, 1 g/L casein hydrolysate, 2 g/L K-glutamate, and 0.335 µmol/L $MgSO_4 \cdot 7H_2O$), which was put in a shaking incubator at 37°C for 3.5 hours. Donor DNA was extracted from 1 mL of an overnight culture of the donor strain growing in LB using the Bacterial and Yeast Genomic DNA Purification Kit from EURx (Gdansk, Poland), which yielded a typical purified DNA concentration ranging from 50 to 150 ng/µL quantified by Nanodrop. 2 µL donor DNA was added to a fresh tube and mixed with 400 µL of competent cells and further incubated for 2 hours at 37°C before plating of 100 µL on LB agar with appropriate antibiotics for selection of transformants, and incubated overnight at 37°C before selection of successful transformants.

Comparative genome analysis

To identify the gene cassettes correlating with pigment production, a pangenome was constructed using Panaroo (25–28) and subsequently used in a pangenome association study (Pan-GWAS) with pigmentation phenotypes in pyseer (29). Due to the low sample size of pigment producers (3 out of 13), we accepted hits with bad χ test flags. The resulting hits (Data set S1) were sorted by *P*-value and effect size (beta) and analyzed manually to identify targets where the association with pigmentation could be confirmed by gene deletion. Genomes were additionally re-annotated using bakta (30) at the default settings for any additional information on the identified genes.

Production of ^{13}C,^{15}N-labeled samples

The *B. subtilis* MB9_B4 GFP and MB9_B4 GFP Δ*yetJ*-PGC strains (Table S1) were cultured in 5 mL of LB, at 37°C, 120 rpm overnight, as pre-culture. 5 mL of labelled MSgg buffer (5 mM K_2HPO_4/KH_2PO_4 pH7, 100 mM MOPS pH7, 0.5% L-glutamic acid $^{13}C_5$ ^{15}N, 0.5% glycerol $^{13}C_3$, 50 µM tryptophan, 50 µM threonine, 2 mM $MgCl_2$ $6H_2O$, 2 µM thiamine, 700 µM $CaCl_2$, 100 µM $MnCl_2$, 50 µM $FeCl_3$, and 1 µM $ZnCl_2$) were then inoculated with pre-culture at 0.1% and cultured at 30°C without agitation for 5 days in order to induce pellicle biofilm formation at the surface. The media culture was collected with a syringe with a needle 21G and centrifuged at 15,000 *g*, 30 minutes, 4°C in order to keep the supernatant. The pellet containing the biofilm was recovered with a spatula and used for spectroscopic analysis.

Solid-state NMR and ATR-FTIR spectroscopies

^{13}C,^{15}N-labeled samples were analyzed by magic-angle spinning solid-state NMR (MAS NMR) spectroscopy. One-dimensional ^{13}C and ^{15}N spectra were recorded on a 600 MHz Bruker magnet equipped with a 4 mm triple resonance probe. An 11 kHz spinning frequency and high-power ^1H decoupling (83 kHz) were used to record 1D insensitive nuclei enhanced by polarization transfer (INEPT) using acquisition times of 20 and 15 ms for ^{13}C and ^{15}N spectra, respectively. Chemical shifts were calibrated according to sodium

trimethylsilylpropanesulfonate. ATR-FTIR spectroscopy was performed on a Nicolet iS50 FTIR spectrometer equipped with a mercury cadmium telluride (MCT) detector cooled with liquid N_2 and an ATR accessory mounted with a germanium crystal (one reflection). Spectra were recorded with 100 scans and a spectral resolution of 2 cm^{-1}.

RESULTS

Culture conditions for pigment production

When testing the antifungal activity of *B. subtilis* isolates on PDA medium (31), we observed that MB9_B4 produces a pigmented compound that diffuses into the agar medium. To get a better understanding of the conditions under which the pigment is produced, we screened pigmentation under a range of culture media (Fig. 1).

In colony biofilm culturing, we were able to reproduce the same pigmentation observed in earlier studies on PDA (Fig. 1A). More interestingly, pigmentation was even stronger on biofilm-inducing media such as MSgg (chemically defined biofilm-promoting medium) and LBgm (LB supplemented with glycerol and manganese that induces biofilm development in *B. subtilis*) compared to PDA. Further testing of LBg (LB supplemented with glycerol) and LBm (LB supplemented with manganese) featured pigmentation only on LBg, although LB itself contains trace amounts of manganese. Lastly, pigmentation was also observed on King's B agar, which is a known inducer of fluorescent pigments in *Pseudomonas spp*.

To probe the importance of certain additives on pigmentation, we performed a medium component analysis in MSgg agar, leaving out each component from the chemically defined medium, which revealed that pigment production did not depend on any single one of the tested components (Fig. 1B).

Pigmentation was observed at 25°C (although delayed by 1 day) and 35°C but was absent at 45°C (Fig. 2A). We noted a variation in coloration, which was clearest on MSgg, with low temperatures resulting in a more "bluish" pigment and higher temperatures leading to a "brownish" coloration. We tested if the variable coloration was a physiological response to the temperature or a chemical change in the pigmentation

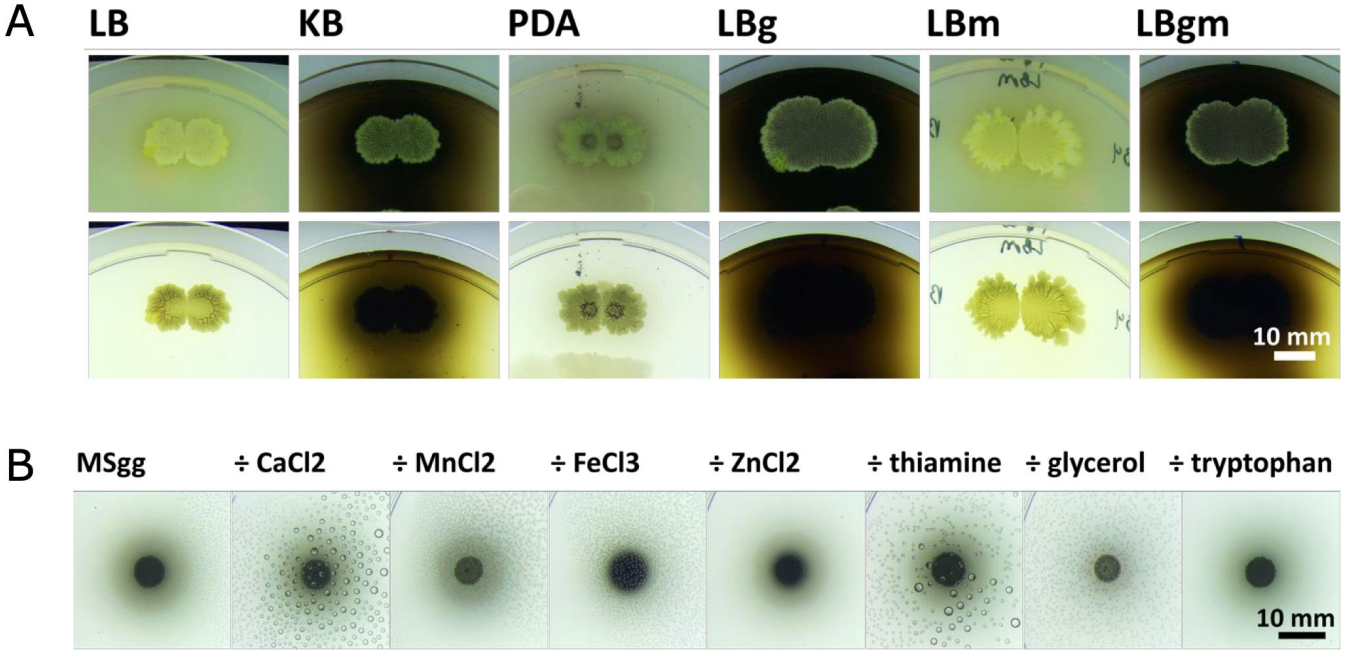

FIG 1 (A) Pigmentation of MB9_B4 on various solid media after 5 days at room temperature imaged with either contrasting light from top (top row) or below (bottom row) the plate, scale bar corresponds to 10 mm. (B) Medium component analysis of MB9_B5 pigmentation on MSgg agar after 5 days at room temperature, showing complete media to the left and each component omitted as noted above. Separate images with a scale bar in the lower left corner corresponding to 10 mm.

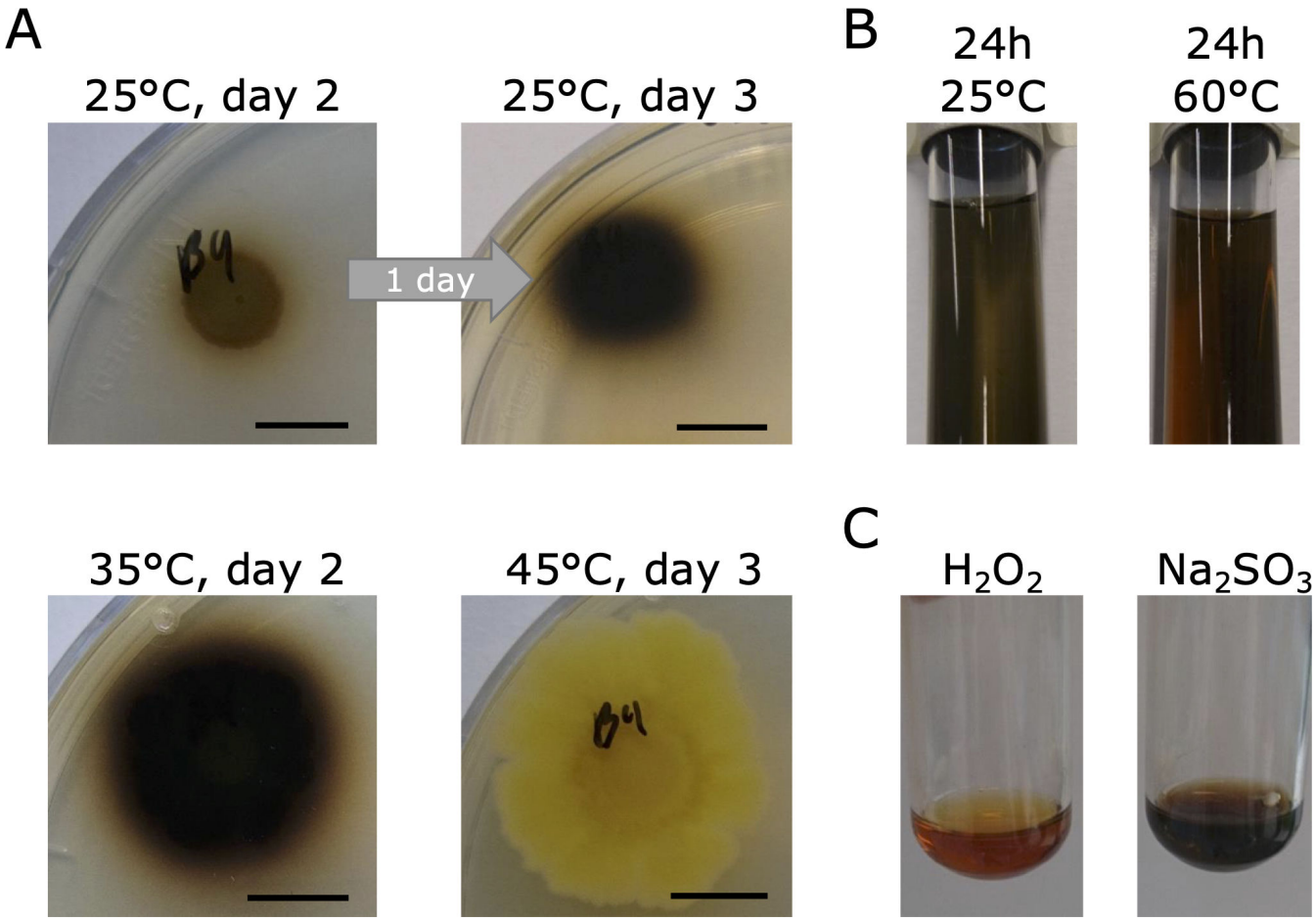

FIG 2 (A) Pigmentation on MSgg agar plates kept for 2 days at 25, 35, and 45°C, respectively; day 3 picture shown for 25°C condition as pigmentation improved considerably. Scale bar in each picture corresponds to 10 mm. (B) Change in coloration of 10× diluted pigment kept for 24 hours at 60°C compared to 25°C. (C) Effect of 24 hour treatment with 1% v/v H_2O_2 and 1% w/v Na_2SO_3.

itself by sterile filtering diluted pigmented medium and incubating at higher temperatures, which resulted in a shift to a brown color indicating a temperature-dependent decay (Fig. 2B).

To check if the change in coloration was due to oxidation, we further tested the effect of hydrogen peroxide (1% v/v H_2O_2) as well as the reducing agent sodium sulfite (1% w/v Na_2SO_3). Oxidizers were able to bleach the pigment, while the reducing agent did seemingly not affect coloration (Fig. 2C). Additionally, the oxidation was irreversible; supplementation of the reducing agent to the oxidized cultures resulted in no further change.

GWAS and confirmation of pigment production BGC by deletion

We screened 13 additional *B. subtilis* isolates from our strain collection (31) for pigmentation on MSgg agar plates to facilitate comparative genomics (Fig. S1). Here we found pigmentation in three out of the 13 isolates (MB9_B4, 63, and 73). Comparing the pigmented isolates with non-pigmented strains revealed the presence of genes in the pigmented strains that are annotated as homologues of the tryptophan biosynthetic operon co-located within a ~17 kbp gene cluster between the *yetJ* and *yetK* loci, which we designated as pigment gene cluster (PGC) (Fig. 3A; Tables 1 and 2). The connection between pigment production and the identified gene cluster was subsequently confirmed by deleting the *B. subtilis* MB9_B4 PGC between *yetJ* and *yetK* using homologous recombination of the upstream and downstream regions present in the *yetJ::kan*^{R}

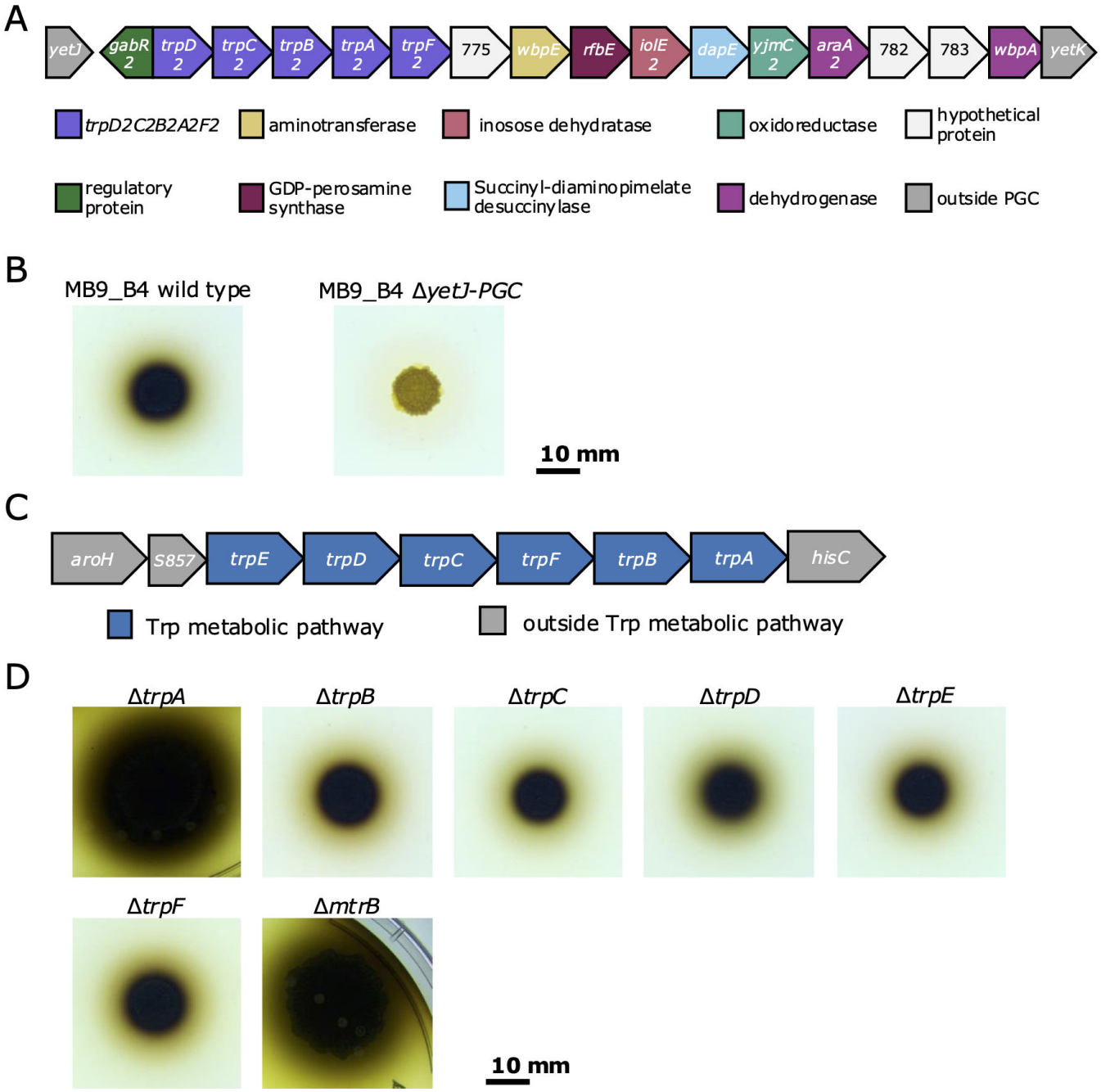

FIG 3 (A) Diagram of the PGC, coloration by product class and presumed function from annotation, additional information on each gene can be found in Tables 1 and 2. (B) Deletion of *yetJ*-PGC results in loss of pigmentation on MSgg at 30°C, image from a 4 day old sample. (C) Diagram of the canonical tryptophan metabolic gene cluster in *B. subtilis* (blue) and surrounding genes for context (grey). (D) Pigmentation is unaffected by the deletion of *mtrB* and *trpABCDEF* biosynthetic genes. Scale bar corresponds to 10 mm for the images in B and C.

and *yetK*::*kan*^{R} mutant strains of the BKK library (BKK07200 and BKK07210, respectively) (Fig. 3B, Δ*yetK* shown in Fig. S2). During integration of the *kan*^{R} antibiotic cassettes, recombination occurs at the upstream and downstream regions of *yetJ* and *yetK* genes of the 168 strains (used for the BKK library), resulting in removal of the PGC in MB9_B4.

Since multiple genes were annotated to encode part of tryptophan biosynthesis, we next tested if the PGC could complement the deletion of the canonical *trpA-F* genes or if pigment production was affected by tryptophan regulatory elements (Fig. 3C, other tested mutants can be found in Fig. S3). Single deletion mutants of the *trpA-F* genes in

TABLE 1 Genes in PGC by their abbreviated name used in Fig. 3A with name and length annotated by PGAP and bakta

Abbreviation	Name (PGAP)	Name (bakta)	Length (PGAP)	Length (bakta)
gabR2	*gabR_4*	JONKOO_20605	1,416	1,416
trpD2	*trpD_1*	JONKOO_20610	1,059	1,059
trpC2	*trpC_1*	*trpC*	783	783
trpB2	*trpB_1*	*trpB*	1,197	1,197
trpA2	*trpA_1*	*trpA*	819	819
trpF2	*trpF_1*	JONKOO_20630	654	654
775	ACGIOGKE_00775	JONKOO_20635	1,029	486
		JONKOO_20640		474
wbpE	*wbpE*	JONKOO_20645	1,152	1,152
rfbE	*rfbE*	JONKOO_20650	1,128	1,128
iolE2	*iolE_1*	JONKOO_20655	870	870
dapE	*dapE*	JONKOO_20660	1,296	1,296
yjmC2	*yjmC_1*	JONKOO_20665	1,104	1,104
araA2	*araA_1*	JONKOO_20670	903	903
782	ACGIOGKE_00782	*yopX*	414	414
783	ACGIOGKE_00783	JONKOO_20680	690	567
wbpA	*wbpA*	JONKOO_20685	1,335	1,335

MB9_B4 were unable to grow on media lacking tryptophan. Likewise, deletion of the *mtrB* gene encoding the RNA-binding attenuation protein of tryptophan biosynthesis did not affect pigmentation.

Blast search using the genes of the PGC revealed an additional 27 genomes or scaffolds carrying a PGC with high identity (Fig. 4). All isolates that carried a PGC belonged to the *B. subtilis* species complex, with the majority being either *B. subtilis* or *Bacillus mojavensis*. The PGC itself was remarkably conserved across most PGC-carrying

TABLE 2 Genes in PGC by their abbreviated name used in Fig. 3A with product annotated by PGAP and bakta

Abbreviation	Product (PGAP)	Product (bakta)
gabR2	HTH-type transcriptional regulatory protein GabR	PLP-dependent aminotransferase family protein
trpD2	Anthranilate phosphoribosyltransferase	Anthranilate phosphoribosyltransferase
trpC2	Indole-3-glycerolphosphate synthase	Indole-3-glycerol phosphate synthase TrpC
trpB2	Tryptophan synthase beta chain	Tryptophan synthase subunit beta
trpA2	Tryptophan synthase alpha chain	Tryptophan synthase subunit alpha
trpF2	N-(5′-phosphoribosyl)anthranilate isomerase	Phosphoribosylanthranilate isomerase
775	Hypothetical protein	Gfo/Idh/MocA family oxidoreductase, frameshift
wbpE	UDP-2-acetamido-2-deoxy-3-oxo-D-glucuronate aminotransferase	DegT/DnrJ/EryC1/StrS family aminotransferase
rfbE	GDP-perosamine synthase	DegT/DnrJ/EryC1/StrS family aminotransferase
iolE2	Inososе dehydratase	Sugar phosphate isomerase/epimerase
dapE	Succinyl-diaminopimelate desuccinylase	M20/M25/M40 family metallo-hydrolase
yjmC2	Putative oxidoreductase YjmC	Ldh family oxidoreductase
araA2	L-arabinose1-dehydrogenase (NAD(*P*)(+))	Gfo/Idh/MocA family oxidoreductase
782	Hypothetical protein	YopX domain-containing protein
783	Hypothetical protein	B3-4 domain-containing protein
wbpA	UDP-N-acetyl-D-glucosamine 6-dehydrogenase	Nucleotide sugar dehydrogenase

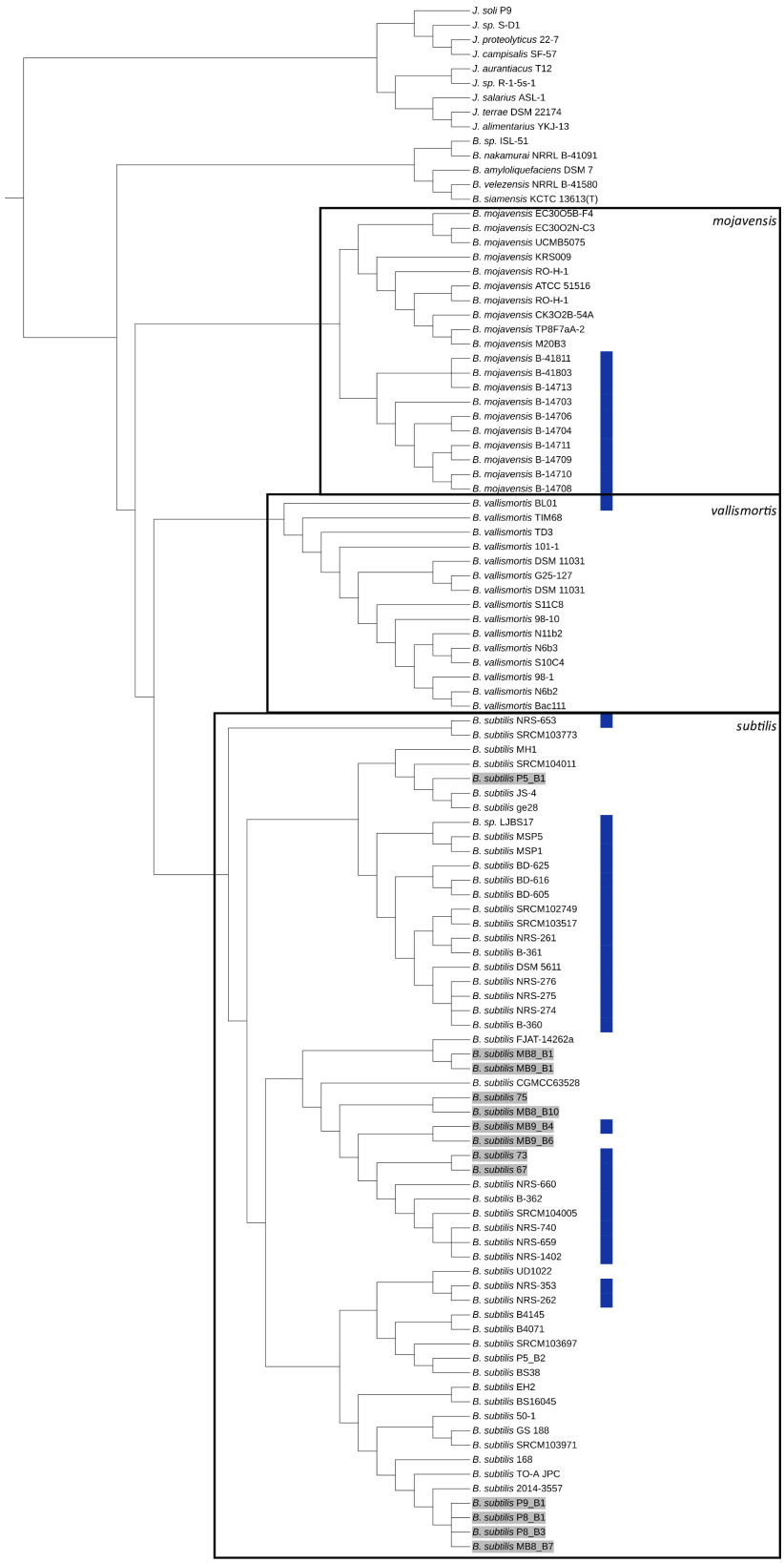

FIG 4 Phylogenetic tree (based on autoMLST) of pigment-carrying genomes and closely related strains, PGC-harboring strains are marked with a blue ribbon. Isolates tested in the experiments are highlighted with a grey background. Figure created using iTOL (32).

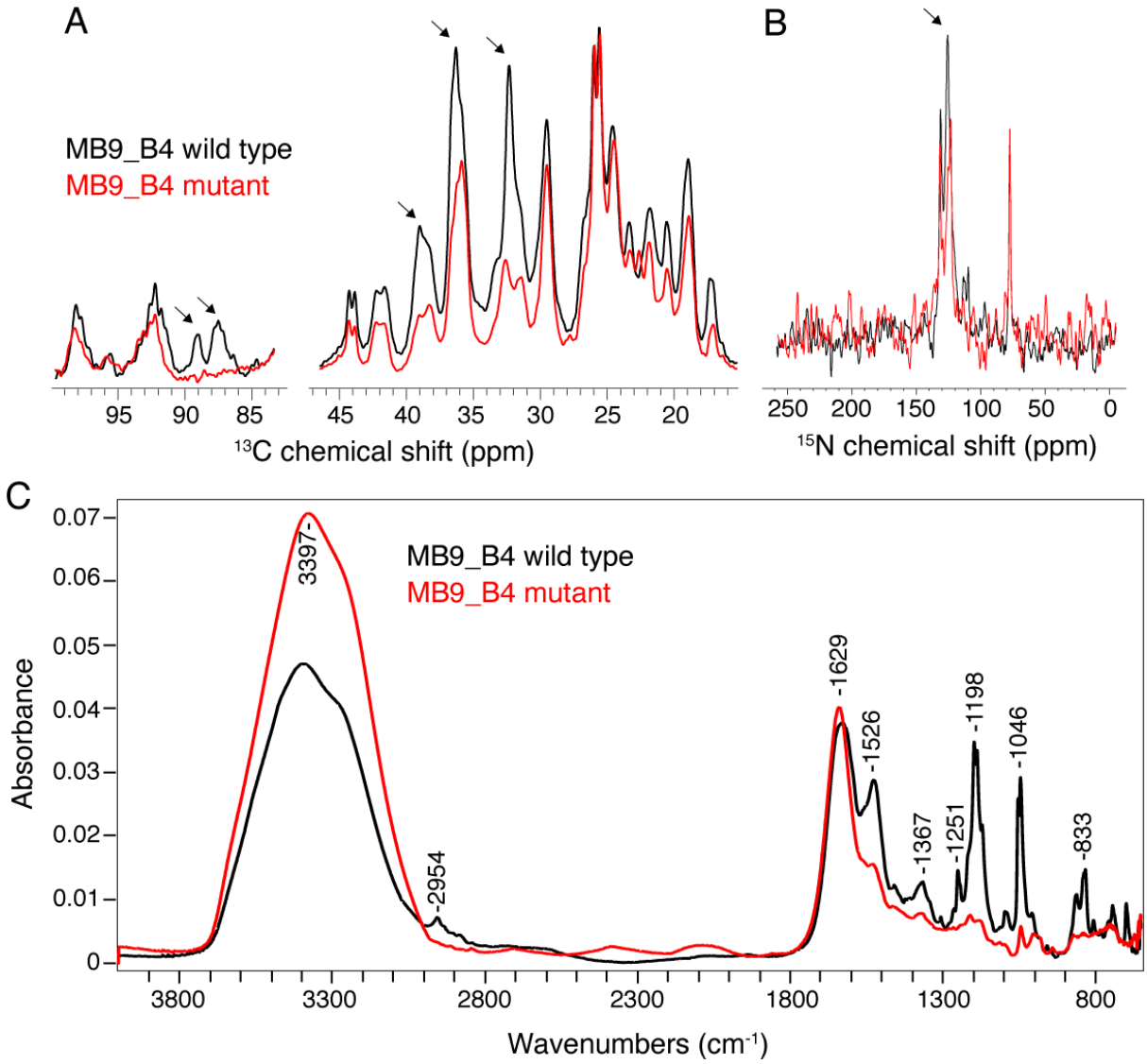

FIG 5 Spectroscopic analysis of ^{13}C,^{15}N-labeled *Bacillus* pellicles for the wild-type MB9_B4 (black lines) and the mutant lacking PGC (MB9_B4 GFP Δ*yetJ*-PGC, red lines). Magic-angle spinning solid-state NMR analysis of pellicles recorded at 600 MHz at a spinning frequency of 11 kHz, using one-dimensional ^{1}H-^{13}C and ^{1}H-^{15}N INEPT spectra, respectively, in panels A and B. (C) Fourier transform infrared absorbance spectra.

isolates (ANI = 100 – 98.19%) except NRS-0660, which was lacking the first ~5,000 bps and NRS-0276, where only fragments of the PGC could be identified. This might be due to incomplete genome sequences, as the genome sequences of NRS-0660 and NRS-0276 are not completed and closed. Corresponding strains were tested for pigmentation on MSgg agar. Here, NRS-0660 was able to produce the pigment, while NRS-0276 exhibited a light purple coloration. Out of the remaining 25 isolates, a further four were unable to produce pigmentation in the assayed conditions, although there were no signs of disruption of the PGC in these.

Structural elucidation by spectroscopy

To get more detailed information about the chemical nature of the pigment, we carried out a spectroscopic analysis of *B. subtilis* MB9_B4 and its PGC-deleted mutant (Fig. 5; Table S2). Bacterial pellicles constitute extraordinarily complex analytical matrices made of many chemical entities (polysaccharides, proteins, amino acids, lipids, and water), rendering their high-resolution analysis challenging. Magic-angle spinning solid-state

NMR (MAS NMR) and ATR-FTIR spectroscopies constitute powerful analytical tools to investigate the molecular composition of complex insoluble biomolecular pigments (33), as demonstrated for melanin samples of different sources (34, 35). We employed one-dimensional ^{1}H-^{13}C and ^{1}H-^{15}N INEPT experiments under magic-angle spinning conditions (11 kHz) on ^{13}C, ^{15}N-labeled samples to obtain the molecular fingerprint of *Bacillus* spp. pellicles. The ^{13}C-detected experiment (Fig. 5A) reveals a typical complex spectroscopic signature encoding for protein, lipids, polysaccharides, and other molecules (36). To guide our NMR signal assignment, we compared our data to previous NMR studies on pigments (37–41). By comparing wild-type and PGC-deleted mutant strains, we detected several spectral areas associated with a substantial reduction of ^{13}C signals encoding for alkyl (e.g., methylene) groups at ~32, 36, and 38 ppm for the PGC-deleted strain. Additionally, in the range of 86–90 ppm, two major signals were absent for the PGC-deleted strain, which could be putatively assigned to –CHOH or –CHOAr. Many microbial pigments are nitrogen-containing molecules, and we probed the structural ^{15}N environment with a ^{1}H-^{15}N INEPT (Fig. 5B). We observed a loss of signal in the mutant at ^{15}N ~125 ppm, a spectral region that usually encodes nitrogen signal of melanin-like pigments (37, 42). Next, we performed ATR-FTIR to identify the chemical bonds by comparing the fingerprints of the same two samples (Fig. 5C). Noticeable differences were observed for absorbance at 2,954 cm^{-1} (aliphatic stretching), 1,200 cm^{-1} (phenolic C-OH stretching), and 1,046 cm^{-1} (C-H aromatic in plane). A strong contribution at 833 cm^{-1} (out-of-plan bending), often observed in melanin-like samples, was also considerably reduced in the mutant sample. Note that a small absorbance difference (~20 cm^{-1}) is observed in our samples (^{13}C-labeled) compared to reported values from the literature obtained on unlabeled samples. Altogether, MAS NMR and ATR-FITR pointed to a pigment containing alkyl, C = C and CHO groups.

DISCUSSION

SMs represent a huge reservoir of untapped potential, with a large majority of antibiotics and therapeutics originating from microbial products. However, while some classes of SMs are studied in detail, it is hypothesized that a large majority of the biosynthetic potential is still relatively unknown. The current tools for the discovery of SMs largely rely on the similarity of the biosynthetic apparatus encoding genes to known gene clusters and are challenged when it comes to finding truly novel SM-related BGCs; thus, examples of novel BGCs serve greatly to increase the available parts of the SM potential.

Here, we identify a melanin-like pigment produced in specific *B. subtilis* isolates. Our test strains exhibited pigmentation on several biofilm-inducing media, with the presence of glycerol and, to a minor degree, manganese being common to most of the highly inducing conditions, which could indicate biofilm induction through the glycerol-manganese pathway observed by Shemesh et al. (22). The importance of glycerol and manganese was further probed in LB and MSgg, where the addition of glycerol but not manganese was found to be essential in LB, while neither was essential for pigmentation in MSgg. The differences observed between LB and MSgg and the generally late timing of production could indicate that pigment production is associated with biofilm and/or sporulation in *B. subtilis,* although this remains to be experimentally elucidated.

Strains featuring the same type of pigmentation in our collection allowed us to identify the responsible BGC, hence dubbed PGC. The 17 kbp PGC consists of 16 putative genes located between the genes *yetJ* and *yetK* and was found present in an additional 27 genomes, some of which have been studied as early as in 1896 under the nomination *B. aterrimus* in reference to their dark pigmentation. Interestingly, several genes in the PGC were annotated as parts of the tryptophan biosynthesis operon, which could suggest an origin from the primary metabolism. However, in our setup, the PGC *trp* versions were unable to complement deletions of the canonical *trp* genes. Thus, if PGC genes originated as duplicates of the canonical *trp* operon, they have been fully repurposed beyond the promiscuity that some enzymes exhibit during neofunctionalization (43). Besides *trp*-like genes, annotations for the PGC included

several aminotransferases, oxidoreductases, one sugar phosphate isomerase/epimerase, one metallo-hydrolase, and one nucleotide sugar dehydrogenase, some of which also seemed to originate from the primary metabolism.

The spectra from MAS NMR and FTIR revealed several peaks characteristic of a melanin-like pigment containing alkyl, C = C, CHO, and –CHOH groups. Melanins are a diverse group of polymeric pigments, which can further be subdivided into several classes depending on coloration and chemical composition: eumelanin, pheomelanin, pyomelanin, allomelanin, neuromelanin, and DHN melanin (44–46). While eumelanin, pheomelanin, pyomelanin, and neuromelanin usually feature L-tyrosine as a precursor (44, 45) and DHN melanin uses either acetyl CoA or malonyl CoA (47), tryptophan has also been described as a possible precursor both synthetically (48) and in the bacterium *Rubrivivax benzoatilyticus* (49). The PGC identified here includes several *trp* genes, which, altogether with the similar solubility and interactions with H_2O_2 noted in reference 49, could entertain the possibility that the pigment identified in *B. subtilis* could be based on tryptophan as a precursor. The putative –CHOH signals around 86–90 ppm additionally could indicate the presence of a sugar moiety in the pigment, which would further match with the presence of a nucleotide sugar dehydrogenase in the PGC. While our spectroscopic study focused on native pellicles, most studies from the literature have been carried out on purified microbial melanin samples (44, 50–52). Therefore, several spectral ambiguities hamper the exact chemical structure determination of the pigment. Further advanced purification protocols, ideally complemented with mass spectrometry analysis, will be required to drastically simplify the pellicle sample and obtain pure pigment for high-resolution structural analysis.

A previous study on *B. aterrimus* from 1946 described a gradual loss of capability to produce the pigment on potato when serially cultured in glucose broth or on glucose-nitrate agar (1 g/L K_2HPO_4, 1 g/L $NaNO_3$, 13 g/L agar, and 10 g/L glucose), which could sometimes be regained by serial culturing on potato again (21). Glucose-nitrate media and potato differ widely in their composition, but such difference could be the key to determining the function of the pigment, as growth on potato likely conveyed an advantage for pigment-producing derivatives and allowed pigmentation to reappear in the strain that had otherwise lost pigment production.

We believe the PGC identified in this study is a valuable addition to our knowledge as an example of a BGC with lowly diverged duplications from the primary metabolism, which could aid our understanding of the evolution of secondary metabolism and improve bioinformatics tools for BGC mining in the future. The pigment itself is likely an example of a bacterial *trp*-melanin for which the biosynthetic genes have not before been described.

ACKNOWLEDGMENTS

We thank Heiko T. Kiesewalter for his observation of the dark pigment in *B. subtilis* isolate 67 and Xinming Xu for her help in creating the phylogenetic tree.

This project was supported by the Danish National Research Foundation (DNRF137) for the Center for Microbial Secondary Metabolites and the Novo Nordisk Foundation for the "Imaging microbial language in biocontrol (IMLiB)" infrastructure grant (NNF19OC0055625). This work has benefited from the Biophysical and Structural Chemistry Platform at IECB, CNRS UAR 3033, and INSERM US001. This work was supported in part by a USDA in-house project (5010-22410-024-000-D).

Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the U.S. Department of Agriculture. The mention of firm names or trade products does not imply that they are endorsed or recommended by the USDA over other firms or similar products not mentioned. USDA is an equal opportunity provider and employer.

AUTHOR AFFILIATIONS

[1]DTU Bioengineering, Technical University of Denmark, Kongens Lyngby, Denmark
[2]Institute of Biology Leiden, Leiden University, Leiden, Netherlands
[3]USDA ARS NCAUR, Peoria, Illinois, USA
[4]University of Bordeaux, CNRS, Inserm, IECB, UAR3033, US01, Pessac, France
[5]University of Bordeaux, CNRS, Bordeaux INP, CBMN, UMR, 5248, IECB, Pessac, France

AUTHOR ORCIDs

Rune Overlund Stannius http://orcid.org/0000-0003-4036-7301
Estelle Morvan http://orcid.org/0000-0003-4071-7885
Mélanie Berbon http://orcid.org/0000-0003-3235-253X
Antoine Loquet http://orcid.org/0000-0001-7176-7813
Ákos T. Kovács http://orcid.org/0000-0002-4465-1636

FUNDING

Funder	Grant(s)	Author(s)
Danish National Research Foundation	DNRF137	Ákos T. Kovács
Novo Nordisk Fonden	NNF19OC0055625	Ákos T. Kovács
Biophysical and Structural Chemistry Platform		Antoine Loquet
U.S. Department of Agriculture	5010-22410-024-000-D	Christopher A. Dunlap

AUTHOR CONTRIBUTIONS

Rune Overlund Stannius, Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft | Christopher A. Dunlap, Formal analysis, Resources, Writing – review and editing | Estelle Morvan, Investigation, Methodology | Mélanie Berbon, Investigation, Methodology | Sophie Lecomte, Investigation, Methodology | Antoine Loquet, Methodology, Resources, Supervision, Writing – review and editing | Ákos T. Kovács, Conceptualization, Funding acquisition, Resources, Supervision, Writing – original draft

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Data set S1 (mSystems00759-25-s0001.xlsx). Pan-GWAS output.
Supplemental material (mSystems00759-25-s0002.pdf). Fig. S1 to S3 and Tables S1 and S2.

Open Peer Review

PEER REVIEW HISTORY (review-history.pdf). An accounting of the reviewer comments and feedback.

REFERENCES

1. Kovács ÁT. 2019. *Bacillus subtilis*. Trends Microbiol 27:724–725. https://doi.org/10.1016/j.tim.2019.03.008
2. Stragier P, Losick R. 1996. Molecular genetics of sporulation in *Bacillus subtilis*. Annu Rev Genet 30:297–341. https://doi.org/10.1146/annurev.genet.30.1.297
3. Errington J. 2003. Regulation of endospore formation in *Bacillus subtilis*. Nat Rev Microbiol 1:117–126. https://doi.org/10.1038/nrmicro750
4. Piggot PJ, Hilbert DW. 2004. Sporulation of *Bacillus subtilis*. Curr Opin Microbiol 7:579–586. https://doi.org/10.1016/j.mib.2004.10.001
5. Duc LH, Fraser PD, Tam NKM, Cutting SM. 2006. Carotenoids present in halotolerant *Bacillus* spore formers. FEMS Microbiol Lett 255:215–224. https://doi.org/10.1111/j.1574-6968.2005.00091.x
6. Khaneja R, Perez-Fons L, Fakhry S, Baccigalupi L, Steiger S, To E, Sandmann G, Dong TC, Ricca E, Fraser PD, Cutting SM. 2010. Carotenoids

found in *Bacillus*. J Appl Microbiol 108:1889–1902. https://doi.org/10.1111/j.1365-2672.2009.04590.x
7. Manzo N, D'Apuzzo E, Coutinho PM, Cutting SM, Henrissat B, Ricca E. 2011. Carbohydrate-active enzymes from pigmented *Bacilli*: a genomic approach to assess carbohydrate utilization and degradation. BMC Microbiol 11:198. https://doi.org/10.1186/1471-2180-11-198
8. Steiger S, Perez-Fons L, Fraser PD, Sandmann G. 2012. Biosynthesis of a novel C_{30} carotenoid in *Bacillus firmus* isolates. J Appl Microbiol 113:888–895. https://doi.org/10.1111/j.1365-2672.2012.05377.x
9. Barnett TA, Hageman JH. 1983. Characterization of a brown pigment from *Bacillus subtilis* cultures. Can J Microbiol 29:309–315. https://doi.org/10.1139/m83-051
10. Charron-Lamoureux V, Haroune L, Pomerleau M, Hall L, Orban F, Leroux J, Rizzi A, Bourassa J-S, Fontaine N, d'Astous ÉV, Dauphin-Ducharme P, Legault CY, Bellenger J-P, Beauregard PB. 2023. Pulcherriminic acid modulates iron availability and protects against oxidative stress during microbial interactions. Nat Commun 14:2536. https://doi.org/10.1038/s41467-023-38222-0
11. Angelini LL, Dos Santos RAC, Fox G, Paruthiyil S, Gozzi K, Shemesh M, Chai Y. 2023. Pulcherrimin protects *Bacillus subtilis* against oxidative stress during biofilm development. NPJ Biofilms Microbiomes 9:50. https://doi.org/10.1038/s41522-023-00418-z
12. Arnaouteli S, Matoz-Fernandez DA, Porter M, Kalamara M, Abbott J, MacPhee CE, Davidson FA, Stanley-Wall NR. 2019. Pulcherrimin formation controls growth arrest of the *Bacillus subtilis* biofilm. Proc Natl Acad Sci U S A 116:13553–13562. https://doi.org/10.1073/pnas.1903982116
13. Ramírez-Rendon D, Passari AK, Ruiz-Villafán B, Rodríguez-Sanoja R, Sánchez S, Demain AL. 2022. Impact of novel microbial secondary metabolites on the pharma industry. Appl Microbiol Biotechnol 106:1855–1878. https://doi.org/10.1007/s00253-022-11821-5
14. Newman DJ, Cragg GM. 2020. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. J Nat Prod 83:770–803. https://doi.org/10.1021/acs.jnatprod.9b01285
15. Kim JH, Lee N, Hwang S, Kim W, Lee Y, Cho S, Palsson BO, Cho B-K. 2021. Discovery of novel secondary metabolites encoded in actinomycete genomes through coculture. J Ind Microbiol Biotechnol 48:kuaa001. https://doi.org/10.1093/jimb/kuaa001
16. Sélem-Mojica N, Aguilar C, Gutiérrez-García K, Martínez-Guerrero CE, Barona-Gómez F. 2019. EvoMining reveals the origin and fate of natural product biosynthetic enzymes. Microb Genomics 5:e000260. https://doi.org/10.1099/mgen.0.000260
17. Caetano-Anollés G, Yafremava LS, Gee H, Caetano-Anollés D, Kim HS, Mittenthal JE. 2009. The origin and evolution of modern metabolism. Int J Biochem Cell Biol 41:285–297. https://doi.org/10.1016/j.biocel.2008.08.022
18. Cruz-Morales P, Kopp JF, Martínez-Guerrero C, Yáñez-Guerra LA, Selem-Mojica N, Ramos-Aboites H, Feldmann J, Barona-Gómez F. 2016. Phylogenomic analysis of natural products biosynthetic gene clusters allows discovery of arseno-organic metabolites in model *Streptomycetes*. Genome Biol Evol 8:1906–1916. https://doi.org/10.1093/gbe/evw125
19. Lehmann KB, Neumann RO. 1896. Atlas und Grundriss der Bakteriologie und Lehrbuch der speziellen bakteriologischen Diagnostik. J. F. Lehmann.
20. Clark FE, Smith NR. 1939. Cultural requirements for the production of black pigments by Bacilli. J Bacteriol 37:277–284. https://doi.org/10.1128/jb.37.3.277-284.1939
21. Smith NR, Smith NR, Gordon RE, Clark FE. 1946. Aerobic mesophilic sporeforming bacteria. U.S. Department of Agriculture.
22. Shemesh M, Chai Y. 2013. A combination of glycerol and manganese promotes biofilm formation in *Bacillus subtilis* via histidine kinase KinD signaling. J Bacteriol 195:2747–2754. https://doi.org/10.1128/JB.00028-13
23. Branda SS, González-Pastor JE, Ben-Yehuda S, Losick R, Kolter R. 2001. Fruiting body formation by *Bacillus subtilis*. Proc Natl Acad Sci U S A 98:11621–11626. https://doi.org/10.1073/pnas.191384198
24. Kunst F, Rapoport G. 1995. Salt stress is an environmental signal affecting degradative enzyme synthesis in *Bacillus subtilis*. J Bacteriol 177:2403–2407. https://doi.org/10.1128/jb.177.9.2403-2407.1995
25. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, Gladstone RA, Lo S, Beaudoin C, Floto RA, Frost SDW, Corander J, Bentley SD, Parkhill J. 2020. Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biol 21:180. https://doi.org/10.1186/s13059-020-02090-4
26. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153
27. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066. https://doi.org/10.1093/nar/gkf436
28. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37:1530–1534. https://doi.org/10.1093/molbev/msaa015
29. Lees JA, Galardini M, Bentley SD, Weiser JN, Corander J. 2018. Pyseer: a comprehensive tool for microbial pangenome-wide association studies. Bioinformatics 34:4310–4312. https://doi.org/10.1093/bioinformatics/bty539
30. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. 2021. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. bioRxiv. https://doi.org/10.1101/2021.09.02.458689
31. Kiesewalter HT, Lozano-Andrade CN, Wibowo M, Strube ML, Maróti G, Snyder D, Jørgensen TS, Larsen TO, Cooper VS, Weber T, Kovács ÁT. 2021. Genomic and chemical diversity of *Bacillus subtilis* secondary metabolites against plant pathogenic fungi. mSystems 6:e00770-20. https://doi.org/10.1128/mSystems.00770-20
32. Letunic I, Bork P. 2024. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res 52:W78–W82. https://doi.org/10.1093/nar/gkae268
33. Song W, Yang H, Liu S, Yu H, Li D, Li P, Xing R. 2023. Melanin: insights into structure, analysis, and biological activities for future development. J Mater Chem B 11:7528–7543. https://doi.org/10.1039/d3tb01132a
34. Zhong J, Frases S, Wang H, Casadevall A, Stark RE. 2008. Following fungal melanin biosynthesis with solid-state NMR: biopolymer molecular structures and possible connections to cell-wall polysaccharides. Biochemistry 47:4701–4710. https://doi.org/10.1021/bi702093r
35. Glass K, Ito S, Wilby PR, Sota T, Nakamura A, Bowers CR, Vinther J, Dutta S, Summons R, Briggs DEG, Wakamatsu K, Simon JD. 2012. Direct chemical evidence for eumelanin pigment from the Jurassic period. Proc Natl Acad Sci U S A 109:10218–10223. https://doi.org/10.1073/pnas.1118448109
36. Reichhardt C, Cegelski L. 2014. Solid-state NMR for bacterial biofilms. Mol Phys 112:887–894. https://doi.org/10.1080/00268976.2013.837983
37. Adhyaru BB, Akhmedov NG, Katritzky AR, Bowers CR. 2003. Solid-state cross-polarization magic angle spinning ^{13}C and ^{15}N NMR characterization of *Sepia* melanin, *Sepia* melanin free acid and *Human hair* melanin in comparison with several model compounds. Magn Reson Chem 41:466–474. https://doi.org/10.1002/mrc.1193
38. Chrissian C, Camacho E, Kelly JE, Wang H, Casadevall A, Stark RE. 2020. Solid-state NMR spectroscopy identifies three classes of lipids in *Cryptococcus neoformans* melanized cell walls and whole fungal cells. J Biol Chem 295:15083–15096. https://doi.org/10.1074/jbc.RA120.015201
39. Chatterjee S, Prados-Rosales R, Frases S, Itin B, Casadevall A, Stark RE. 2012. Using solid-state NMR to monitor the molecular consequences of *Cryptococcus neoformans* melanization with different catecholamine precursors. Biochemistry 51:6080–6088. https://doi.org/10.1021/bi300325m
40. Hervé M, Hirschinger J, Granger P, Gilard P, Deflandre A, Goetz N. 1994. A ^{13}C solid-state NMR study of the structure and auto-oxidation process of natural and synthetic melanins. Biochim Biophys Acta 1204:19–27. https://doi.org/10.1016/0167-4838(94)90027-2
41. Duff GA, Roberts JE, Foster N. 1988. Analysis of the structure of synthetic and natural melanins by solid-phase NMR. Biochemistry 27:7112–7116. https://doi.org/10.1021/bi00418a067
42. Chatterjee S, Prados-Rosales R, Tan S, Itin B, Casadevall A, Stark RE. 2014. Demonstration of a common indole-based aromatic core in natural and synthetic eumelanins by solid-state NMR. Org Biomol Chem 12:6730–6736. https://doi.org/10.1039/c4ob01066c
43. Noda-García L, Juárez-Vázquez AL, Ávila-Arcos MC, Verduzco-Castro EA, Montero-Morán G, Gaytán P, Carrillo-Tripp M, Barona-Gómez F. 2015. Insights into the evolution of enzyme substrate promiscuity after the discovery of $(\beta\alpha)_8$ isomerase evolutionary intermediates from a diverse metagenome. BMC Evol Biol 15:107. https://doi.org/10.1186/s12862-015-0378-1
44. El-Naggar N-A, El-Ewasy SM. 2017. Bioproduction, characterization, anticancer and antioxidant activities of extracellular melanin pigment

produced by newly isolated microbial cell factories *Streptomyces glaucescens* NEAE-H. Sci Rep 7:42129. https://doi.org/10.1038/srep42129
45. Plonka PM, Grabacka M. 2006. Melanin synthesis in microorganisms - biotechnological and medical aspects. Acta Biochim Pol 53:429–443. https://doi.org/10.18388/abp.2006_3314
46. d'Ischia M, Wakamatsu K, Napolitano A, Briganti S, Garcia - Borron J, Kovacs D, Meredith P, Pezzella A, Picardo M, Sarna T, Simon JD, Ito S. 2013. Melanins and melanogenesis: methods, standards, protocols. Pigment Cell Melanoma Res 26:616–633. https://doi.org/10.1111/pcmr.12121
47. Eisenman HC, Casadevall A. 2012. Synthesis and assembly of fungal melanin. Appl Microbiol Biotechnol 93:931–940. https://doi.org/10.1007/s00253-011-3777-2
48. Vèkey K, Tamás J, Somogyi A, Bertazzo A, Costa C, Allegri G, Seraglia R, Traldi P. 1992. Studies on structure characterization of tryptophan melanin: comparison between filament and curie - point pyrolysis gas chromatography/mass spectrometry. Org Mass Spectrom 27:1216–1219. https://doi.org/10.1002/oms.1210271111
49. Ahmad S, Mohammed M, Mekala LP, Chintalapati S, Chintalapati VR. 2020. Tryptophan, a non-canonical melanin precursor: new L-tryptophan based melanin production by *Rubrivivax benzoatilyticus* JA2. Sci Rep 10:8925. https://doi.org/10.1038/s41598-020-65803-6
50. Sajjan S, Kulkarni G, Yaligara V, Kyoung L, Karegoudar TB. 2010. Purification and physiochemical characterization of melanin pigment from *Klebsiella* sp. GSK. J Microbiol Biotechnol 20:1513–1520. https://doi.org/10.4014/jmb.1002.02006
51. Correa N, Covarrubias C, Rodas PI, Hermosilla G, Olate VR, Valdés C, Meyer W, Magne F, Tapia CV. 2017. Differential antifungal activity of human and cryptococcal melanins with structural discrepancies. Front Microbiol 8:1292. https://doi.org/10.3389/fmicb.2017.01292
52. Liporagi-Lopes LC, Chrissian C, Kacirani A, Camacho E, Stark RE, Casadevall A. 2024. New insights into the melanin structure of *Lomentospora prolificans*. bioRxiv:2024.11.01.621558. https://doi.org/10.1101/2024.11.01.621558