Phenogenomic characterization of a newly domesticated and novel species from the genus *Verrucosispora*

Sarah J. Kennedy, [a] Celine Grace F. Atkinson, [a]* Brooke R. Tomlinson [a], Lauren Hammond, [a] Prahathees Eswara, [a] Bill J. Baker, [b] Lindsey N. Shaw[a]#

[a]Department of Cell Biology, Microbiology and Molecular Biology, University of South Florida, Tampa, Florida, USA

[b]Department of Chemistry, University of South Florida, Tampa, Florida, USA

Running Head: Domestication of *Verrucosispora sioxanthis* sp. nov.

#Address correspondence to Lindsey N. Shaw, shaw@usf.edu.

*Present address: Celine Grace F. Atkinson, Department of Pediatrics, Morsani College of Medicine, University of South Florida, Tampa, Florida, USA

Abstract word count: 200

Main body of text word count: 5150

Abstract

The concept of bacterial dark matter stems from our inability to culture most microbes and represents a fundamental hole in our knowledge of microbial diversity. Herein we present the domestication of such an organism: a previously uncultured, novel species from the rare-Actinomycetes genus *Verrucosispora*. Although initial recovery took >4 months, isolation of phenotypically distinct, domesticated generations occurred within weeks. Two isolates were subjected to phenogenomic analyses, revealing domestication correlated with enhanced growth rates in nutrient-rich media, but diminished capacity to metabolize diverse amino acids. This is seemingly mediated by genomic decay through the pseudogenization of amino acids metabolism genes. Conversely, later generational strains had enhanced spore germination rates, potentially through the reversion of a sporulation-associated kinase from pseudogene to true gene status. We observed that our most wild-type isolate had the greatest potential for antibacterial activity, which correlated with extensive mutational attrition of biosynthetic gene clusters in domesticated strains. Comparative analyses revealed wholesale genomic reordering in strains, with widespread SNP, indel and pseudogene mutations observed. We hypothesize that domestication of this previously unculturable organism resulted from the shedding of genomic flexibility required for life in a dynamic marine environment, parsing out genetic redundancy to allow for a newfound cultivable amenability.

Importance

The majority of environmental bacteria cannot be cultured within the laboratory. Understanding why only certain environmental isolates can be recovered is key to unlocking the abundant microbial dark matter that is widespread on our planet. In this study we present not only the culturing but domestication of just such an organism. Although initial recovery took >4 months, we were able to isolate distinct, sub-passaged offspring from the originating colony within mere weeks. A phenotypic and genotypic analysis of our generational strains revealed that adaptation to life in the lab occurred as a result of wholesale mutational changes. These permitted an enhanced ability for growth in nutrient rich media, but came at the expense of reduced genomic flexibility. We suggest that without dynamic natural environmental stressors our domesticated strains effectively underwent genomic decay as they adapted to unchallenging conditions experienced in the laboratory.

Introduction

The Great Plate Count Anomaly has confounded microbiologists since the 1800s and the introduction of solid growth media (1). The first recorded instance was in 1932 by Razumov, who was unable to equate viable plate counts with the microscopic enumeration of environmental isolates (2). Advances in next generation sequencing and metagenomics in the modern era have only confirmed this culturable discrepancy, leaving the longest unresolved question in microbiology unanswered. Recent analyses of genomic data produce a bleak picture, highlighting that the current rate of characterization at 600 new species per year will require more than 1,000 years to complete (3). The knowledge from this microbial “dark matter” has untold and innumerable consequences. Unlike the hypothetical understanding of dark matter in astrophysics, concrete microbial dark matter discoveries divulge the uncultured majority as a dominant factor within the universe (4). Technological advancements reveals that this uncharacterized diversity also encodes the potential to produce countless bioactive compounds that must be investigated (5), both for bacterial taxonomy and to potentially restock our chemical arsenal in a post-antibiotic age (6).

While the post-antibiotic age is a relatively new threat, humans have been using bacterial natural products since ancient Mesopotamia (7). More recently, however, we have been able to optimize bacterial biosynthetic potential for myriad applications. This domestication, or taming, is analogous to domestication of farm animals and crops. Human selection has caused “wild type” organisms to evolve away from their highly

variable and complex natural environments: without dynamic natural environmental stressors necessitating fitness traits, domesticated organisms effectively undergo genomic decay as they adapt to static, unchallenging conditions experienced in the laboratory (8).

During our explorations of cryptic microbial dark matter, we have previously reported recovery and whole genome sequence of *Verrucosispora* sp. CWR15 isolated from a Gulf of Mexico sponge. Although markedly understudied, the *Verrucosispora* genus has attracted some attention due to the demonstratable bioactive potential of its secondary metabolites. Since the taxon was established in 1998, there have been discoveries of novel antibiotic (9), antitumor (10), anti-influenza A (11), and anti-HIV activities (12, 13). Despite this, the majority marine-derived genus is among the “rare actinomycetes” reticent to laboratory culturing and domestication (14).

In this study at hand we report the domestication of *Verrucosispora* sp. CWR15 by sequencing laboratory sub-passaged, phenotypically distinct generations and their functional characterization. Genomic exploration reveals laboratory-induced genetic changes in key functions like metabolism, environmental information processing, antibacterial gene clusters and antimicrobial resistances. This results in a domesticated, previously unculturable organism that appears to be a novel species, herein named *Verrucosispora sioxanthis*. This domestication and speciation serve as an example of how to decrease the microbial dark matter knowledge gap with the aim of restocking the antibiotic medicine cupboard in a post-antibiotic age.

Methods

Serial passaging: Three distinctive generations of *Verrucosispora sioxanthis* CWR15 were obtained via liquid sub-culturing. Each generation was grown for 7d in identical baffled flasks with 60mL tryptic soy broth (TSB; 30g/L) containing sucrose (20% (v/v) solution of filter sterilized 50% (w/v) sucrose) at 30°C at 210rpm. Serial passaging was accomplished by withdrawing 1mL of cultures and inoculating them into fresh TSB/sucrose flasks every 7d. Each generation was preserved at -80°C in 20% glycerol in Instant Ocean (IO; 36g/L) solution.

Spore solution generation: Individual colonies were grown into 5mL TSB containing sterile glass beads and sucrose for 7-14d at 210rpm and 30°C. Cultures were plated on TSA supplemented with sucrose and placed in a humidified incubator at 30°C for 7-14d until lawns established. Spores were scraped from agar plates into filter-sterilized water (15). Cellular debris was filtered using autoclaved coffee filters, and solutions were incubated at 55°C for 10min to kill vegetative cells prior to experimentation.

Conditional growth assay: TSB, Luria-Bertani Broth (LB; 10g/L tryptone, 5g/L yeast extract, 10g/L NaCl), AMM (10g/L starch, 4g/L yeast extract, 2g/L peptone, 36g/L IO) and International *Streptomyces* Project 2 (ISP-2; 10g/L malt extract, 4g/L glucose, 4g/L yeast extract, 36g/L IO) were used for these assays in 96-well plates sealed with parafilm and surrounded with a sterile moat of water to minimize evaporation over the 14d+ experiment. Five replicates of each media (75µL) were used with/without sucrose.

126 Freshly processed spore solution was inoculated to each well, to a final volume of
127 100µL. An identical set-up containing a UV-sterilized glass bead (0.5mm) in each well
128 was also used. Plates were incubated in a humidified shaker at 30°C and 210rpm, with
129 growth monitored via OD_{600} every 24h using a Synergy 2 Microplate reader .
130
131 **Methylene blue assay:** The absorption and desorption of germinating spore solutions
132 was monitored as described previously (16). Five 24-well plates were inoculated with
133 TSB and spore solution, along with three sterile glass beads (3mm). Sealed plates were
134 incubated in a humidified shaker at 30°C and 210rpm. Methylene blue absorption was
135 performed within the growth plate to limit transferal loss, at 245rpm. Complete
136 methylene blue desorption employed two washes with 250mM HCl. Desorption was
137 determined by reading the OD_{660} of supernatants. The exact concentrations of
138 methylene blue absorbed and desorbed was determined using a previously established
139 absorption coefficient (16).
140
141 **Genomic DNA extraction and sequencing:** Genomic DNA was extracted from 7d
142 TSB/sucrose/beads cultures incubated as described above. Cultures were mechanically
143 disrupted with sterile glass beads before undergoing phenol chloroform DNA extraction.
144 Quality was monitored with a Thermo Scientific NanoDrop® ND-1000 UV-Vis
145 Spectrophotometer. An Ion Torrent PGM Hi-Q View OT2 Kit was used to generate 200-
146 300bp library fragments with ~100ng of input DNA. Ion Plus Fragment Library Kit
147 Adapters and manufacturer's protocols were used to ligate adapters to fragments, and
148 Ion Xpress barcodes were attached to allow multiple samples on a single chip.

Purification was performed with Agencourt AMPure XP Reagent according to manufacturer's protocol, and library size selection was performed with an E-Gel SizeSelect Agarose Gel. Amplification was performed using an Ion Plus Fragment Library Kit, purified with Agencourt AMPure XP Reagent, and analyzed with an Agilent High Sensitivity DNA Kit and an Agilent 2100 Bioanalyzer instrument. Barcoded libraries for each generation were pooled, and an Ion OneTouch 2 instrument was used to prepare template-positive Ion PGM Hi-Q View Ion Sphere Particles (ISPs). ISPs were enriched using an Ion OneTouch ES instrument prior to sequencing using an Ion 318 v2 chip (Ion Torrent).

***In silico* prediction of species novelty:** The hybrid assembly for generation 3 (H-G3, used for scaffolding purposes) has been described previously (17). Determination of dDDH and taxonomy was achieved by submitting the H-G3 genome through the Type (Strain) Genome Server (TYGS), available at https://tygs.dsmz.de/ (18). The determination of closely related type strains (19-22), pairwise comparison of genome sequences (22), phylogenetic inferences (23-25), and type-based species and subspecies clustering (26) are described in (18). The DSMZ-generated phylogenetic tree was exported to the Interactive Tree Of Life (iTOL) for visualization (27). Average nucleotide identity (ANI) (28) and average amino acid identity (AAI) (29) of H-G3 was repeated herein following the taxonomic reorganization of *Verrucosispora* and *Micromonospora*.

Domestication elucidation bioinformatics: For the genetic investigation of domestication, Ion Torrent PGM reads were trimmed and processed using "map reads to contigs" of the previously published H-G3 (17) within CLC Genomics Workbench (CLCGW, v20.0.2), default parameters. The resulting contigs for each generation were aligned to *Verrucosispora* CNZ293 using progressiveMauve (30) (default settings, except allowing use seed families option enabled) to generate a universal positional identifying number for the generations to be plotted to. Generational genomic accessions are available within **Data availability** section. The new genomes were aligned (G1:G3) in progressiveMauve using previously described parameters. The mapping coverage data exported from CLCGW generated a coverage histogram for G1. Biosynthetic gene clusters (BGCs) were identified using bacterial antiSMASH v6.0.0 (31) and annotated CDS were obtained from CLCGW. For G1-G3 comparisons, homologous genes were determined by reciprocal best BLASTn hit. Gene tracks were generated from reference genomes and gene annotations were extracted from each gene track using CLCGW. Extracted annotations were used in a reciprocal BLASTn (32) with then-named *Verrucosispora maris* (currently *Micromonospora maris*) and *Verrucosispora* sp. NA02020 genome annotations as local databases, and vice versa, using default BLASTn settings with the flag "-max_target_seqs 1" to restrict results to the lowest E-value for homologous determination of KEGG ontological groupings.

To explore domestication impact, extracted CDS nucleotide (nt) and translated amino acid (aa) sequences from G1 and G3 genomes were aligned using Clustal Omega (v1.2.3) (33), default parameters except wrapping options were enabled to prevent

wrapping (--wrap=10,000 for nt, --wrap=5,000 for aa). This was done in a pairwise manner to compare equivalent proteins, unified across generations by annotation locus tags.

G1 and G3 genome assemblies were submitted to antiSMASH v6.0.0 (31) with "relaxed" strictness and all extra features enabled to identify BGCs. The previously described SNP and indel analyses were applied to BGCs, with a modification to prevent Clustal Omega wrapping (--wrap=1,000,000). The GC-content and contig average mapping coverage depth were determined with CLCGW v21.0.3.

Fluorescence microscopy: Fluorescence microscopy was performed as previously described (34). Cultures were grown in TSB with sucrose and glass beads (30°C and 210rpm) for 7d. Aliquots (10µL) of each culture were diluted in 200µL of PBS, and then 100µl was mixed with BODIPY FL dye (1µg/ml) and incubated for 10min at room temperature. FM4-64 (1µg/ml) was added and 5µl was spotted onto a glass bottom dish (MatTek) and covered with a 1% agarose pad. Microscopy was performed using a GE Applied Precision DeltaVision Elite deconvolution fluorescence microscope equipped with a Photometrics CoolSnap HQ2 camera. Seventeen planes were acquired for each image, each 200nm apart. Files were deconvolved using the *softWoRx* software v6.5.2.

Amino acid utilization: Methodology involving nitrogen-deficient basal media was used as described previously (35). Briefly, 7d liquid cultures were pelleted and washed to

remove growth media. Duplicate inocula were prepared as described (35) for duplicate un-supplemented basal media control plates and supplemented plates. Plates were sealed with parafilm and placed in a humidified incubator at 30°C for 21d prior to scoring results according to literature. Colonies were counted and reported by comparison to basal media for baseline analyses. Sources of amino acids (0.1% w/v final concentration) were: β-alanine (Calbiochem); L-arginine, proline, serine (Acros Organics); L-glutamic acid (Sigma-Aldrich); glycine (Fisher).

Secondary metabolite extraction: Three biological replicates of *Verrucosispora sioxanthis* were grown in 60mL baffled flasks in TSB, ISP-2, AMM or 2YT (10g/L yeast extract, 16g/L tryptone, and 5g/L NaCl) (36) at 30°C and 210rpm. Following 21d incubation, 30mL ethyl acetate (EtOAc) was added to cultures before being returned to the incubator for 1h. The resulting extract was filtered through commercial grade coffee filters to remove bacterial biomass, before being left to settle for 24h. The organic layers were transferred to pre-weighed scintillation vials and dried overnight via an airline. An additional 30mL EtOAc was added to original culture media for a second extraction 24h later. Following two extractions, the final dried extract was solvated in 100% dimethyl sulfoxide (DMSO) at 5mg/mL.

Antimicrobial testing: Overnight cultures of ESKAPE pathogens were utilized in antibacterial screens as previously described (37). Plates were incubated at 37°C and 245rpm overnight. Inhibitory activity was determined via OD_{600} readings within a Synergy 2 microplate reader.

Data availability: *Verrucosispora sioxanthis* generational genomic contigs are deposited in NCBI under the BioProject PRJNA734818. Generation 1 (G1) is available within the BioSample SAMN19546627, and generation 3 (G3) is available within SAMN19546628. The G1 Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under accession JAHMAD000000000, G3 has been deposited at DDBJ/ENA/GenBank under accession JAHMAC000000000.

Results

Cultivation and domestication of a novel isolate of the *Verrucosispora* genus: We have previously detailed the isolation and genome sequencing of a novel isolate from the genus *Verrucosispora*, *Verrucosispora* sp. CWR15, isolated from a Gulf of Mexico sponge (17). Given the understudied nature of the *Verrucosispora* genus, and the rarity with which they are cultured, we present herein a detailed phenogenomic exploration of the generational domestication of this organism. Following initial isolation of *Verrucosispora* sp. CWR15 we observed that its continued culturing within the lab proved difficult. Very often the organism failed to disperse in broth, and media preparations would evaporate or desiccate during extended (7+ days) incubations. To explore this, we initially tried different standard laboratory media (TSB and LB) without the addition of the environmental sponge extract that *Verrucosispora* sp. CWR15 was initially cultured on (**Figures 1A** and **1B**). In so doing, we noted minimal growth in TSB and LB alone that was only improved upon inclusion of both sucrose and glass beads, alongside extended culture times. Towards this latter point, sucrose is known to incite faster/non-aggregative growth in similarly hyphal-growing organisms such as *Streptomyces* spp. (38), whilst sterile glass beads physically disrupt biomass in liquid media and mechanically prevent aggregation. We next tried more specialized media typically used to culture organisms such as *Verrucosispora* spp. (39-41). Here we found that AMM and ISP-2 (**Figure 1C and 1D**) media were much better suited for rapid biomass accumulation, as evidenced by the 3-5 day increase in optical density for AMM, and for ISP-2 after the 6-day time point. Unfortunately, although these media

produced more robust growth, they are not amenable to downstream experimentation due to component solubility and instability, particularly during long growth windows. Indeed, such instability was evidenced by the drastic and sometimes unpredictable growth patterns observed. Thus, all downstream experimentation was performed with TSB + sucrose and beads, yielding reproducible and steady growth.

Serial sub-culturing yields generationally distinct isolates with disparate growth rates: Using supplemented TSB, we were able to sub-passage *Verrucosispora* sp. CWR15. This domestication, via gradual introduction to laboratory culturing, yielded visually distinct generations over time that grew to different biomasses with varying efficiencies. Strikingly, the first (G1), third (G3), and fifth (G5) generations were notable as being the most phenotypically distinct. Initial differences were observed in dispersal, biomass, and rate of growth between each generation. This is remarkable as: *(1)* no record exists of this rarely cultured taxon displaying such laboratory-attenuation; *(2)* such drastic phenotypic changes are typically described for bacteria spanning years not days (42); and *(3)* the elucidation of laboratory-attenuation mechanisms could provide the basis for understanding why this rarely cultured taxon resists traditional culturing. To avoid complications from colony aggregation, we first measured growth rates of our generational isolates using a modified methylene blue absorption protocol. We observed that the three generations desorb methylene blue at different concentrations based on their biomass (**Figure 2**). Although G1 may follow the same overall trends as the other isolates, it never accumulates the same biomass as its offspring generations. Indeed, even after four days, G1 has significantly less biomass than G2 and G3.

Contrastingly, G3 has the most biomass at all timepoints, although G3 and G5 eventually converge. Together, these initial growth trends suggest that the most "wild-type" organism, G1, struggles to achieve the biomass of the more lab-attenuated G3 and G5. Such growth reticence of earlier generations likely contributes to the taxon's culturing infrequency.

Whole genome sequencing of generational isolates: With generationally distinct strains in hand, we next performed whole-genome sequencing on our two most phenotypically distinct isolates (G1 and G3). This was done to shed light on the rapid domestication observed, and to glean insight as to why this taxon resists culturing. Although we have previously published a hybridized Nanopore and Illumina genome for H-G3 (17), herein we performed IonTorrent-based sequencing for G1 and G3 for ease of comparison. These were individually re-mapped to the hybridized G3 (H-G3) genome to provide standardized generational comparisons (**Figure 3**)**.** Since the published H-G3 genome consisted of 35 contigs in no order, comparable universal position numbers were derived using progressiveMauve to align contigs to the single contig scaffold reference, *Verrucosispora* CNZ293. To account for differences between G1 and G3 genomes, all analyses examined only genes present within both. While G1 has a total of 5,018 CDS annotated, G3 only has 4,762 CDS. Amongst these, 4,420 genes were present in both strains (88% of G1 CDS), presenting only a minor limitation to these analyses. Generational mapping coverage depths and statistics are shown in **Supplemental Table S1**; whilst the bioinformatics pipeline used herein is in **Supplemental Figure S1.**

Domestication induces SNPs, indels and widespread pseudogene alterations throughout the genome: To explore domestication mutations, we subtracted Mauve-identified SNPs from the Clustal-identified mutations, allowing indel identification. In so doing, it appears that the majority of genetic changes between our isolates is not due to SNPs (285/4,420 genes, ~6.4%), but instead is indels (2,570/4,420 genes, ~58.1%). When calculating mutation frequencies per gene (accounting for varying gene lengths) we noted the SNP frequencies ranged from 0-3.5%, while indel frequencies ranged from 0-54%. To link KEGG ontological groups with CDS via BLASTn, we next pooled G1 and G3 CDS and align them with the genomes of *Verrucosispora maris* (now *Micromonospora maris*) and *Verrucosispora* NA02020. While these references only confer KEGG ontology to 1,181/4,420 G1+G3 genes (~26.7%), it is not due to a lack of homology. ANI values are 84.52% for *Verrucosispora/Micromonospora maris* and 84.91% for *Verrucosispora* NA02020, indicating similarity. Indeed, only 6 total CDS were not matched in the reciprocal BLASTn's. The most significant contribution to the lack of KEGG ontology is that ~40% and ~41% of *Verrucosispora/Micromonospora maris* and *Verrucosispora* NA02020 CDS, respectively, are not associated with a KEGG ontological group. Another point to note is that indels caused significant generational DNA dissimilarity, as determined via Clustal Omega. Here, indel-identified genes ranged from 46-99.97% DNA similarity: 1,778/4,420 genes (~40.2%) contained $\leq$ 10 indels; 546/4,420 genes (~12.4%) contained 11-50 indels, 210/4,420 genes (~4.8%) contained 51-199 indels; and 36/4,420 genes (~0.8%) contained 201-1611 indels. Overall, only 271/4,420 genes (~6.1%) contained both SNP and indel mutations.

Since pseudogenes present within the un-sequenced originating sponge strain cannot be identified, the presence or absence of extraneous stop codons in G1 cannot be determined. Instead, we explored the change of pseudogene status from G1 to G3 in the context of laboratory attenuation. Upon analysis, we observed that 598/4,420 genes (13.5%) changed pseudogene status from G1 to G3. Of these genes, 191/598 genes (~31.9%) changed from pseudogene status in G1 to non-pseudogene status in G3.

Pseudogene alteration in amino acid metabolism genes results in measurable phenotypic outcome: The largest KEGG ontological group within the references (and therefore G1 and G3) belongs to metabolism, due to the overrepresentation of metabolism within available annotations. Strikingly, numerous amino acid metabolism genes were identified as undergoing mutations in our strains that changed pseudogene status (**Supplemental Table S2)**. These included genes that specify proteins required for the utilizations of arginine, glutamine, proline, and serine. G1 contained six non-pseudogenes and three pseudogenes, whilst the G3 pseudogene profile flipped to majority pseudogene status (six pseudogenes and three non-pseudogenes); all resulting from indels rather than SNPs. To determine if these events had phenotypic outcome, the amino acid utilization profile of G1 and G3 were evaluated. Additionally, although G5 was not subject to whole genome analysis, it was also included to provide context and insight to continued domestication. We determined that G1 grew in all tested conditions, including basal/nitrogen-deficient media (**Figure 4**). In comparison, G3 and G5 were incapable of growth without amino acid supplementation and failed to

362 demonstrate G1-comparable biomass even in the presence of diverse amino acids. Of
363 note, although G1 grew in all conditions, the preferred nitrogen source appeared to be
364 arginine, which also elicited growth for the other generations as well, although not to the
365 same extent. G3 grew on a larger variety of sources (all but glycine) compared to G5,
366 which was only able to grow on arginine, glutamic acid, and glycine.
367
368 **Domestication results in enhanced spore germination rates***:* As domesticated
369 generations accumulate greater biomass independent of amino acid metabolism, we
370 next focused on growth from spore solutions. While the under-annotated nature of the
371 taxon is a limitation, several sporulation-associated genes were identified. One in
372 particular, *cotH*, a kinase regulator implicated in spore-coat assembly and germination
373 (43) notably underwent domestication-induced pseudogene alteration. Strikingly, this
374 gene had only 80.95% nucleotide similarity between G1 and G3, due to ~19.0% indels
375 that cause change from pseudogene to non-pseudogene upon further domestication
376 (**Supplemental Table S2)**. To explore this observation further, spore germination
377 assays were performed with each generational strain (**Figure 5**). We observed that G1
378 experienced delayed and minimal germination, and subsequent growth, as visualized by
379 a single suspended orange colony. Contrastingly, G3 and G5 germinate and begin
380 rapidly growing at around 14 days, with optical densities far exceeding those of the
381 essentially dormant G1. It is noteworthy that domestication-induced indel accumulation
382 within *cotH* results in G3 encoding a complete, non-pseudogene protein. This alteration
383 could be indicative of a broad selection pressure placed on sporulation and germination
384 genes, as functional CotH is required for efficient germination of *Bacillus subtilis* spores

385 (43). Furthermore, CotH is required for outer spore coat structure and integrity (44). A
386 G1 *cotH* pseudogene could thus contribute to decreased heat resistance and
387 subsequent minimal growth of heat-treated spores.
388
389 **Generational strains display differential aggregation phenotypes:** Next, to explore
390 macroscopic aggregative growth phenotypes, fluorescence microscopy with membrane
391 and cell wall staining was performed (**Figure 6**). We observed that G1 displayed a
392 compact, non-diffuse phenotype that was not present within the more filamentous,
393 elongated G3. G5 was observed to have a phenotype similar to G1, however with more
394 filaments branching from core aggregated cells. We thereby classified G5 to have a
395 phenotype averaging both G1 and G3 dispersal. Overall, these images suggest that G1
396 growth is more aggregative than the more domesticated generations. This could explain
397 the limited biomass accumulation of this strain, and its general failure to disperse.
398 Indeed, the overlapping filaments demonstrate a propensity to grow inwards rather than
399 outwards, causing eventual growth limitation and subsequently decreased biomass.
400
401 **Domestication limits the antibacterial capacity of *Verrucosispora* sp. CWR15:**
402 Given that the genus has contributed novel chemical structures in the past, we next
403 investigated the antibacterial potential of our isolates. We began by analyzing BGCs
404 using antiSMASH for H-G3, identifying 18 clusters, including those for NRPS,
405 T1PKS/T2PKS/T3PKS, siderophores, and bacteriocins. This low number is likely a
406 result of the limited taxonomic investigations and annotations of the genus. Indeed,
407 ~72% of the BGCs do not exceed >50% similarity within available sequences, and 5

clusters returned no significant similarity scores at all (**Supplemental Table S3**). These 5 included syntheses of a lasso peptide, NAAGN, two separate terpenes, and a class I lanthipeptide.

To explore antimicrobial potential, G1 was grown in a variety of media prior to crude extraction of secondary metabolites. These extracts were assayed against a panel of ESKAPE pathogens (**Supplemental Table S4**) at a concentration of 250μg/mL (**Figure 7A**). Differential antibacterial activities were observed between culture conditions, with ISP-2, AMM, and 2YT demonstrating minimal inhibitory effects. TSB conditions, however, elicited profound antibacterial activity against *E. faecium.* This inhibition (~91%) is notable for a crude extract, as no fractionation or purification was undertaken to strengthen activity. Next, all three generations were examined for antibacterial production in TSB. While all showed activity, G1 exhibited the greatest anti-*E. faecium* effects (**Figure 7B**), readily exceeding that of G3 (~80%) and G5 (~67%). This finding is fascinating as we demonstrate that the most wild-type-like strain (G1) has the greatest capacity for antibacterial effects, a phenotype that is lost with increasing laboratory domestication. Such a finding has significant potential ramifications for environmental-microbe natural-products drug-discovery.

The PGM G1 and G3 genomes were uploaded to antiSMASH for investigation of the differential generational bioactivity. The passage of G1 to G3 induced both SNPs and indels in BGCs, however, as with the rest of the genome, the majority of mutations were indels. These mutations contributed to antiSMASH predicting three clusters (producing

xanthoferrin, formicamycins A-M, and isorenieratene) in G1 that had no equivalent in G3 (**Supplemental Table S5**). Conversely, 7/18 BGCs noted above for H-G3 were absent from G1, again due to domestication mutations (**Supplemental Table S6**).

We next undertook a dose-response study with G1 (**Figure 7C**), noting that the observed anti-*E. faecium* activity persists even as the tested concentration decreases. The inhibitory effects of G1 crude extracts continued at >50% until only after they were diluted to a concentration of 100μg/mL. Such activity is indeed noteworthy, as, to our knowledge, no anti-*E. faecium* activity has ever been reported within *Verrucosispora* spp.

***Verrucosispora* sp. CWR15 is a novel species:** As a final step, using data from our genomic analyses, we created a 16S rRNA phylogenetic tree (**Figure 8**), comparing our novel isolate to other members of the genus and the closely related *Micromonospora*. This revealed that *Verrucosispora* sp. CWR15 is most similar to *Micromonospora sediminimaris,* previously recognized as *Verrucosispora sediminis*. Due to the taxon undergoing arrangement with *Micromonospor*a, we also predicted the digital DNA-DNA hybridization (dDDH) to determine novelty of the organism (**Supplemental Table S7**). While taxonomy of the genus may be in flux as *Verrucosispora* spp. are increasingly being acknowledged as heterotypic synonyms to *Micromonospora,* the highest dDDH values of 63.9, 46.3, 61.1% (d0, d4, d6 respectively) do not exceed the accepted threshold of 70%, indicating that the isolate is taxonomically distinct. Furthermore, the 92.04% ANI and 94.1% AAI to *Micromonospora sediminimaris* fall within the threshold

for speciation distinction. Collectively, these genomic and *in silico* assessments support the notion that our isolate is a novel species, which we herein term *Verrucosispora sioxanthis* based on the sioxanthin pigment that bestows the organism with its vibrant orange appearance.

Discussion

Bacterial domestication although underexplored, is not unheard of. It is known that endosymbionts accrue significant pseudogenes upon liberation from their eukaryotic hosts (45-48). Indeed, pseudogene numbers within transitioning endosymbiotic bacteria can reach into the thousands, often accounting for over half of the genome (47, 48). It is also known that gene inactivation through pseudogene accumulation is correlated with reduced selective pressures on redundant genes (49). This is synonymous with laboratory attenuation in our study. As the previously sponge-associated *Verrucosispora sioxanthis* was passaged through high-nutrient, stable laboratory conditions, the organism no longer required the genetic flexibility that allowed it to respond to the host, environment, and surrounding sponge microbiome. Thus, we propose that pseudogenization effectively allowed the shedding of its flexible genome to adapt to a low-stress laboratory environment, parsing out genetic redundancy in an otherwise stable environment.

It is important to acknowledge that the only *Verrucosispora* colony recovered from our sponges was on low-nutrient media supplemented with originating sponge extract. While extended incubations are common for recovering organisms from marine or otherwise oligotrophic environs (50), our incubation time (>4 months) is on the more extreme end of the spectrum. We suggest that this organism needed both extended incubation and trace host sponge factors to accumulate biomass as it approached the laboratory-attenuated genetic profile of G1. We observed that passaging, or

481 domesticating, the organism allowed for quicker growth following initial adaptation to the
482 laboratory. Although each generation was given exactly seven days between passages
483 whilst maintaining all incubation parameters, the generational endpoint biomasses are
484 incomparable. The only variable that could explain these discrepancies is thus the
485 domestication mutations accumulated during laboratory adaptation.
486
487 The domestication documented herein is marked by genome-wide alterations. Indeed,
488 the extent of change in pseudogene status observed in our genomic comparisons was
489 quite striking. This clearly indicates a selection pressure for activating/inactivating
490 mutations in domesticated strains of *Verrucosispora sioxanthis*. This ability to either
491 produce a fully functional protein or introduce early stop codons is in line with the
492 genomic decay present in former endosymbionts that shed their eukaryotic hosts. As
493 facultative endosymbionts transition away from their hosts and host-specific selective
494 pressures, genome decay through gene inactivation is observed, mainly through the
495 usage of pseudogenes and truncated (and thus inactive) gene products (51).
496
497 To investigate the phenogenomic effects of domestication, biomass and metabolism
498 assays were performed to supplement *in silico* study. Regarding amino acid
499 metabolism, pseudogene-mediated flexibility seemingly contributed to *Verrucosispora*
500 *sioxanthis'* ability to adapt and flourish within non-native conditions. The amino acid
501 metabolism genes identified all underwent indel-induced pseudogene alterations in G3
502 resulting in less amino acid metabolic ability and an overall pseudogene-dominant
503 profile. This is in direct contrast to G1 that had the greatest amino acid metabolic

capacity and an overall non-pseudogene profile for those implicated genes. G1's non-pseudogene profile for amino acid metabolism benefits the more wild-type organism; the native oligotrophic conditions of a marine endosymbiont require genetic flexibility to maintain nutrient requirements in a dynamic system. Without the selective pressure of nutrient limitations, lab-attenuated G3 underwent genomic decay that led to a decrease capacity for amino acid consumption. This is in line with the theory that genes not conferring fitness are more susceptible to inactivation or otherwise deletional mutations (52).

In addition to amino acid metabolism, we also investigated domestication-impacted secondary metabolism. Since metabolism was the most-affected ontological group, it would follow that both primary and secondary metabolic pathways would be impacted. Indeed, we demonstrate that domestication impacts the antibacterial activity of crude fermentation extracts. The anti-*E. faecium* activity observed decreased from G1 to G3 and G5. Genetically, G1 is obviously most similar to the originating sponge isolate. Environmental isolates, particularly marine sponge endosymbionts, must rely upon their chemical defenses to compete and fend off predation (53), frequently producing some form of antibacterial activity. The domestication of such isolates can lead to genome decay, in which strains lose these defense mechanisms as they become laboratory attenuated (54). This noted decrease in antibacterial potential is corroborated by Anti-SMASH analysis, where indel-driven mutations result in widely variable predicted BGC content between strains.

Notably, while less domesticated generations retain a greater ability to metabolize amino acids, this does not directly contribute to growth rate as G1 was consistently slower at accumulating biomass. This could be explained, in part, by spore germination abilities. In G1, the *cotH* kinase regulator exists as a pseudogene which may result in impaired germination of it spores, a phenotype observed in our assays. In support of this, it is known that CotH influences spore germination in *B. subtilis* (55). Contrastingly, G3 accumulated indels reverting this gene to non-pseudogene status. A fully-functional CotH could allow G3 the ability to germinate, and propagate more rapidly as vegetative cells, ultimately explaining its less spore-dense fluorescence images under microscopy.

While not every previously unculturable bacterium will behave as *Verrucosispora sioxanthis*, we propose that this study is as a model for examining genome-wide effects of domestication. Taken together, the identification of SNPs, indels, and pseudogenes reveals the genetic impacts of domestication. Any of these variables alone would not allow substantial elucidation of the genomic changes required for domestication. Indeed, given that we observed SNPs accounting for 7% of genetic changes, and indels for 58%, mere genomic comparisons alone are insufficient. It is the coupling of genome-derived hypotheses, with phenotypic investigation herein, that demonstrates the holistic value of a combined comparison. Collectively, this generates new insight into the process of domesticating hard-to-tame environmental isolates recalcitrant to laboratory culture. Consequently, we bring light to the subject of microbial dark matter and create the potential to harness such novel phenogenomic information for untold future biochemical applications.

Acknowledgements: This study was supported by grants AI124458 and AI157506 (both LNS) and AI154992 (BJB and LNS) from the National Institute of Allergy and Infectious Diseases.

References

1. Winterberg H. 1898. Zur Methodik der Bakterienzahlung. Zeitschr f Hyg 29:75-93.

2. Staley JT, Konopka A. 1985. Measurement of in Situ Activities of Nonphotosynthetic Microorganisms in Aquatic and Terrestrial Habitats. Annual Review of Microbiology 39:321-346.

3. Yarza P, Yilmaz P, Pruesse E, Glockner FO, Ludwig W, Schleifer KH, Whitman WB, Euzeby J, Amann R, Rossello-Mora R. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat Rev Microbiol 12:635-45.

4. Bernard G, Pathmanathan JS, Lannes R, Lopez P, Bapteste E. 2018. Microbial Dark Matter Investigations: How Microbial Studies Transform Biological Knowledge and Empirically Sketch a Logic of Scientific Discovery. Genome biology and evolution 10:707-715.

5. Doroghazi JR, Albright JC, Goering AW, Ju KS, Haines RR, Tchalukov KA, Labeda DP, Kelleher NL, Metcalf WW. 2014. A roadmap for natural product discovery based on large-scale genomics and metabolomics. Nat Chem Biol 10:963-8.

6. Milshteyn A, Schneider Jessica S, Brady Sean F. 2014. Mining the Metabiome: Identifying Novel Natural Products from Microbial Communities. Chemistry & Biology 21:1211-1223.

7. Pham JV, Yilma MA, Feliz A, Majid MT, Maffetone N, Walker JR, Kim E, Cho HJ, Reynolds JM, Song MC, Park SR, Yoon YJ. 2019. A Review of the Microbial

578 Production of Bioactive Natural Products and Biologics. Frontiers in Microbiology
579 10.
580 8. Gibbons JG, Rinker DC. 2015. The genomics of microbial domestication in the
581 fermented food environment. Current opinion in genetics & development 35:1-8.
582 9. Keller S, Nicholson G, Drahl C, Sorensen E, Fiedler HP, Sussmuth RD. 2007.
583 Abyssomicins G and H and atrop-Abyssomicin C from the Marine Verrucosispora
584 Strain AB-18-032, p 391. JAPAN ANTIBIOTICS RES ASSOC, Japan.
585 10. Igarashi Y, Yu L, Miyanaga S, Fukuda T, Saitoh N, Sakurai H, Saiki I, Alonso-
586 Vega P, Trujillo ME. 2010. Abyssomicin I, a modified polycyclic polyketide from
587 Streptomyces sp. CHI39. J Nat Prod 73:1943-6.
588 11. Zhang J, Li B, Qin Y, Karthik L, Zhu G, Hou C, Jiang L, Liu M, Ye X, Liu M,
589 Hsiang T, Dai H, Zhang L, Liu X. 2020. A new abyssomicin polyketide with anti-
590 influenza A virus activity from a marine-derived Verrucosispora sp. MS100137.
591 Appl Microbiol Biotechnol 104:1533-1543.
592 12. Leon B, Navarro G, Dickey BJ, Stepan G, Tsai A, Jones GS, Morales ME,
593 Barnes T, Ahmadyar S, Tsiang M, Geleziunas R, Cihlar T, Pagratis N, Tian Y, Yu
594 H, Linington RG. 2015. Abyssomicin 2 reactivates latent HIV-1 by a PKC- and
595 HDAC-independent mechanism. Org Lett 17:262-5.
596 13. Song Y, Li Q, Qin F, Sun C, Liang H, Wei X, Wong N-K, Ye L, Zhang Y, Shao M,
597 Ju J. 2017. Neoabyssomicins A–C, polycyclic macrolactones from the deep-sea
598 derived Streptomyces koyangensis SCSIO 5802. Tetrahedron 73:5366-5372.

14. Subramani R, Sipkema D. 2019. Marine Rare Actinomycetes: A Promising Source of Structurally Diverse and Unique Novel Natural Products. Marine Drugs 17:249.

15. Shepherd MD, Kharel MK, Bosserman MA, Rohr J. 2010. Laboratory maintenance of Streptomyces species. Current protocols in microbiology Chapter 10:Unit-10E.1.

16. Fischer M, Sawers RG. 2013. A Universally Applicable and Rapid Method for Measuring the Growth of Streptomyces and Other Filamentous Microorganisms by Methylene Blue Adsorption-Desorption. Applied and Environmental Microbiology 79:4499-4502.

17. Kennedy SJ, Cella E, Jubair M, Azarian T, Baker BJ, Shaw LN. 2020. Draft Genome Sequence of Verrucosispora sp. Strain CWR15, Isolated from a Gulf of Mexico Sponge. Microbiology Resource Announcements 9:e00176-20.

18. Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nature Communications 10:2182.

19. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biology 17:132.

20. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100-8.

21. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

22. Meier-Kolthoff JP, Auch AF, Klenk HP, Goker M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14:60.

23. Lefort V, Desper R, Gascuel O. 2015. FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program. Molecular Biology and Evolution 32:2798-2800.

24. Farris JS. 1972. Estimating Phylogenetic Trees from Distance Matrices. The American Naturalist 106:645-668.

25. Kreft Ł, Botzki A, Coppens F, Vandepoele K, Van Bel M. 2017. PhyD3: a phylogenetic tree viewer with extended phyloXML support for functional genomics data visualization. Bioinformatics 33:2946-2947.

26. Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V, Fiebig A, Rohde C, Rohde M, Fartmann B, Goodwin LA, Chertkov O, Reddy T, Pati A, Ivanova NN, Markowitz V, Kyrpides NC, Woyke T, Goker M, Klenk HP. 2014. Complete genome sequence of DSM 30083(T), the type strain (U5/41(T)) of Escherichia coli, and a proposal for delineating subspecies in microbial taxonomy. Stand Genomic Sci 9:2.

642 27. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for
643 phylogenetic tree display and annotation. Nucleic Acids Res
644 doi:10.1093/nar/gkab301.
645 28. Yoon SH, Ha SM, Lim J, Kwon S, Chun J. 2017. A large-scale evaluation of
646 algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek
647 110:1281-1286.
648 29. Medlar AJ, Törönen P, Holm L. 2018. AAI-profiler: fast proteome-wide
649 exploratory analysis reveals taxonomic identity, misclassification and
650 contamination. Nucleic Acids Res 46:W479-w485.
651 30. Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome
652 alignment with gene gain, loss and rearrangement. PLoS One 5:e11147.
653 31. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber
654 T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining
655 pipeline. Nucleic Acids Research 47:W81-W87.
656 32. Tomlinson BR, Malof ME, Shaw LN. 2020. A Global Transcriptomic Analysis of
657 Staphylococcus aureus Biofilm Formation Across Diverse
658 Clonal Lineages. bioRxiv doi:10.1101/2020.12.17.423160:2020.12.17.423160.
659 33. Sievers F, Higgins DG. 2018. Clustal Omega for making accurate alignments of
660 many protein sequences. Protein Science 27:135-145.
661 34. Brzozowski RS, White ML, Eswara PJ. 2019. Live-Cell Fluorescence Microscopy
662 to Investigate Subcellular Protein Localization and Cell Morphology Changes in
663 Bacteria. J Vis Exp doi:10.3791/59905.

664 35. Williams ST, Goodfellow M, Alderson G, Wellington EM, Sneath PH, Sackin MJ.
665 1983. Numerical classification of Streptomyces and related genera. J Gen
666 Microbiol 129:1743-813.
667 36. Yim SS, Johns NI, Park J, Gomes AL, McBee RM, Richardson M, Ronda C,
668 Chen SP, Garenne D, Noireaux V, Wang HH. 2019. Multiplex transcriptional
669 characterizations across diverse bacterial species using cell-free systems.
670 Molecular systems biology 15:e8875-e8875.
671 37. Fleeman R, LaVoi TM, Santos RG, Morales A, Nefzi A, Welmaker GS, Medina-
672 Franco JL, Giulianotti MA, Houghten RA, Shaw LN. 2015. Combinatorial
673 Libraries As a Tool for the Discovery of Novel, Broad-Spectrum Antibacterial
674 Agents Targeting the ESKAPE Pathogens. Journal of Medicinal Chemistry
675 58:3340-3355.
676 38. Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA. 2000. Practical
677 Streptomyces genetics. The John Innes Foundation, Norwich.
678 39. RHEIMS H, SCHUMANN P, ROHDE M, STACKEBRANDT E. 1998.
679 Verrucosispora gifhornensis gen. nov., sp. nov., a new member of the
680 actinobacterial family Micromonosporaceae. International Journal of Systematic
681 and Evolutionary Microbiology 48:1119-1127.
682 40. Xi L, Zhang L, Ruan J, Huang Y. 2012. Description of Verrucosispora qiuiae sp.
683 nov., isolated from mangrove swamp sediment, and emended description of the
684 genus Verrucosispora. International Journal of Systematic and Evolutionary
685 Microbiology 62:1564-1569.

686 41. Supong K, Suriyachadkun C, Suwanborirux K, Pittayakhajonwut P, Thawai C.
687 2013. Verrucosispora andamanensis sp. nov., isolated from a marine sponge.
688 International Journal of Systematic and Evolutionary Microbiology 63:3970-3974.
689 42. McLoon AL, Guttenplan SB, Kearns DB, Kolter R, Losick R. 2011. Tracing the
690 Domestication of a Biofilm-Forming Bacterium. Journal of Bacteriology 193:2027-
691 2034.
692 43. Nguyen KB, Sreelatha A, Durrant ES, Lopez-Garrido J, Muszewska A,
693 Dudkiewicz M, Grynberg M, Yee S, Pogliano K, Tomchick DR, Pawłowski K,
694 Dixon JE, Tagliabracci VS. 2016. Phosphorylation of spore coat proteins by a
695 family of atypical protein kinases. Proceedings of the National Academy of
696 Sciences 113:E3482-E3491.
697 44. Freitas C, Plannic J, Isticato R, Pelosi A, Zilhão R, Serrano M, Baccigalupi L,
698 Ricca E, Elsholz AKW, Losick R, O Henriques A. 2020. A protein phosphorylation
699 module patterns the Bacillus subtilis spore outer coat. Molecular microbiology
700 114:934-951.
701 45. Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Pontén T, Alsmark
702 UC, Podowski RM, Näslund AK, Eriksson AS, Winkler HH, Kurland CG. 1998.
703 The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
704 Nature 396:133-40.
705 46. Cho NH, Kim HR, Lee JH, Kim SY, Kim J, Cha S, Kim SY, Darby AC, Fuxelius
706 HH, Yin J, Kim JH, Kim J, Lee SJ, Koh YS, Jang WJ, Park KH, Andersson SG,
707 Choi MS, Kim IS. 2007. The Orientia tsutsugamushi genome reveals massive

708 proliferation of conjugative type IV secretion system and host-cell interaction
709 genes. Proc Natl Acad Sci U S A 104:7981-6.
710 47. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honoré
711 N, Garnier T, Churcher C, Harris D, Mungall K, Basham D, Brown D,
712 Chillingworth T, Connor R, Davies RM, Devlin K, Duthoy S, Feltwell T, Fraser A,
713 Hamlin N, Holroyd S, Hornsby T, Jagels K, Lacroix C, Maclean J, Moule S,
714 Murphy L, Oliver K, Quail MA, Rajandream MA, Rutherford KM, Rutter S, Seeger
715 K, Simon S, Simmonds M, Skelton J, Squares R, Squares S, Stevens K, Taylor
716 K, Whitehead S, Woodward JR, Barrell BG. 2001. Massive gene decay in the
717 leprosy bacillus. Nature 409:1007-11.
718 48. Toh H, Weiss BL, Perkin SA, Yamashita A, Oshima K, Hattori M, Aksoy S. 2006.
719 Massive genome erosion and functional adaptations provide insights into the
720 symbiotic lifestyle of Sodalis glossinidius in the tsetse host. Genome Res 16:149-
721 56.
722 49. Goodhead I, Blow F, Brownridge P, Hughes M, Kenny J, Krishna R, McLean L,
723 Pongchaikul P, Beynon R, Darby AC. 2020. Large-scale and significant
724 expression from pseudogenes in Sodalis glossinidius - a facultative bacterial
725 endosymbiont. Microbial genomics 6:e000285.
726 50. Bodor A, Bounedjoum N, Vincze GE, Erdeiné Kis Á, Laczi K, Bende G, Szilágyi
727 Á, Kovács T, Perei K, Rákhely G. 2020. Challenges of unculturable bacteria:
728 environmental perspectives. Reviews in Environmental Science and
729 Bio/Technology 19:1-22.

730 51. Lamelas A, Gosalbes MJ, Manzano-Marín A, Peretó J, Moya A, Latorre A. 2011.
731 Serratia symbiotica from the aphid Cinara cedri: a missing link from facultative to
732 obligate insect endosymbiont. PLoS genetics 7:e1002357-e1002357.
733 52. Ochman H, Davalos LM. 2006. The Nature and Dynamics of Bacterial Genomes.
734 Science 311:1730.
735 53. Matobole RM, van Zyl LJ, Parker-Nance S, Davies-Coleman MT, Trindade M.
736 2017.
737 Antibacterial Activities of Bacteria Isolated from the Marine Sponges Isodictya co
738 mpressa and Higginsia bidentifera Collected from Algoa Bay, South Africa. Mar
739 Drugs 15.
740 54. Steensels J, Gallone B, Voordeckers K, Verstrepen KJ. 2019. Domestication of
741 Industrial Microbes. Current Biology 29:R381-R393.
742 55. Isticato R, Sirec T, Giglio R, Baccigalupi L, Rusciano G, Pesce G, Zito G, Sasso
743 A, De Felice M, Ricca E. 2013. Flexibility of the Prograamme of Spore Coat
744 Formation in Bacillus subtilis: Bypass of CotE Requirement by Over-Production
745 of CotH. PLOS ONE 8:e74949.
746
747

Supplemental Figure Legends

Supplemental Table S1: G1 and G3 PGM reads mapped to H-G3 contigs. The IonTorrent PGM-derived G1 and G3 reads were mapped to the hybrid H-G3 assembly, with average coverage, and length of resulting genome shown. Given that this analysis was driven by IonTorrent sequencing, which carries an inherent technological limitation from homopolymeric tracts and sequencing reporting fidelity, we chose to analyze generational mapping coverage depths. Despite the >70% GC-rich nature of the genome potentially exacerbating the homopolymeric tract limitation, G1's average mapping coverage depth amongst all contigs was 30.2X. This demonstrates that the downstream analysis of this data is not influenced by homopolymer tract limitations, as coverage throughout the genome either approaches or exceeds the minimum coverage threshold for analysis of 30X depth (SR1). Moreover, G3's average coverage mapping depth was 38.2X. To further validate the IonTorrent data, we compared mapping statistics for G3 to the published H-G3 scaffold across the three sequencing platforms used (IonTorrent PGM, Illumina MiSeq, and Oxford Nanpore MinION). As expected, Illumina MiSeq generated the longest genome with the greatest coverage depth, least read mapping variants, and most reads matched to the genome. However, surprisingly, the IonTorrent PGM data overall surpassed the Oxford Nanopore MinION statistics: IonTorrent produced a longer genome with more reads mapped and greater depth coverage. Bioinformatic precedence states that identification of genomic mutations requires sufficient depth coverage to exclude false determinations (SR2). Since the IonTorrent PGM depth coverage scored higher than the newer technology of the

MinION and exceeded 30X, basing our domestication analysis on the PGM sequencing is fundamentally sound

Supplemental Figure S1: Domestication analysis bioinformatics pipeline. *Verrucosispora sioxanthis* was passaged to generation 3 prior to Nanopore MinION and Illumina MiSeq sequencing, Unicycler hybridization, PGAP annotation, and Mauve arrangement of contigs (Hybrid Generation 3 genome, H-G3) (**A**). IonTorrent PGM reads for generations 1 and 3 were individually mapped to H-G3 and annotated from that scaffold (**B**), prior to Mauve-based identification of SNPs, antiSMASH-based identification of biosynthesis, and KEGG-determined ontologies following reciprocal BLASTn with reference strains. Extracted CDS, both nucleotide (nt) and amino acid (aa) sequences were input into ClustalO for generational comparison alignment (**C**).

Supplemental Table S2: Sporulation and amino acid metabolism genes identified as not maintaining pseudogene status upon domestication. One sporulation gene (top, grey shading) *cotH* and nine amino acid metabolism genes (bottom, unshaded) were identified as not maintaining pseudogene status upon passaging generation 1 to generation 3. From left to right: locus, annotation, role, ClustalO-determined generational DNA similarity (%), Mauve-identified SNP frequency (%), indel frequency (%), number of extra stop codons and pseudogene status in generation 1, number of extra stop codons and pseudogene status in generation 3, and if the pseudogene status is maintained upon passaging. The ten genes are ranked top-to-bottom based on increasing generational DNA similarity.

Supplemental Table S3: Biosynthetic potential of *Verrucosispora sioxanthis*. Bacterial antiSMASH version 6.0.0 was used with the generation 3 hybrid assembly (H-G3; GenBank-accessible SAIY00000000). Biosynthetic clusters with no similarity to known clusters are shaded.

Supplemental Table S4: Multi-drug Resistant ESKAPE Pathogens. The resistance profiles for the six ESKAPE pathogens are abbreviated as follows: Gentamycin (Gen), Penicillin G (Pen G), Tetracycline (Tet), Daptomycin (Dap), Vancomycin (Van), Linezolid (Lin), Ampicillin (Amp), Azithromycin (Azit), Chloramphenicol (Chlor), Clindamycin (Clin), Cethromycin (Cet), Erythromycin (Ery), Penicillin (Pen), Tetracycline (Tet), Ciprofloxacin (Cipro), Imipenem (Imi), Ampicillin/Sulbactum (Amp/Sulb), Amikacin (Amik), Aztreonam (Azt), Cefepime (Cfp), Cefetazodime (Cfz), Levofloxacin (Levo), Tobramycin (Tob), Polymyxin B (Poly B).

Supplemental Table S5: Biosynthetic potential of *Verrucosispora sioxanthis* G1. Bacterial antiSMASH version 6.0.0 was used with the G1 genome. Biosynthetic clusters with no equivalent to G3 are shaded. The frequencies of SNPs and Indels (%) was calculated when a comparable G3 biosynthetic gene cluster was present. CLC Genomics Workbench version 21.0.3 determined the average mapping coverage depth of the BGC-containing contigs and BGC G+C content.

Supplemental Table S6: Biosynthetic potential of *Verrucosispora sioxanthis* G3. Bacterial antiSMASH version 6.0.0 was used with the G3 genome. Biosynthetic clusters with no equivalent to G1 are shaded. The frequencies of SNPs and Indels (%) was calculated when a comparable G1 biosynthetic gene cluster was present. CLC Genomics Workbench version 21.0.3 determined the average mapping coverage depth of the BGC-containing contigs and BGC G+C content.

Supplemental Table S7: Predicted *Verrucosispora sioxanthis* DNA-DNA hybridization and G+C content differences within taxon. The Type (Strain) Genome Server https://tygs.dsmz.de/ was used to generate pairwise comparisons to the subject strains (leftmost column, accession provided). The resulting digital DNA-DNA hybridization (dDDH) values based on gene content similarity (d0 and d6) and sequence identity (d4) are presented with the confidence intervals (C.I.) and G+C content differences.

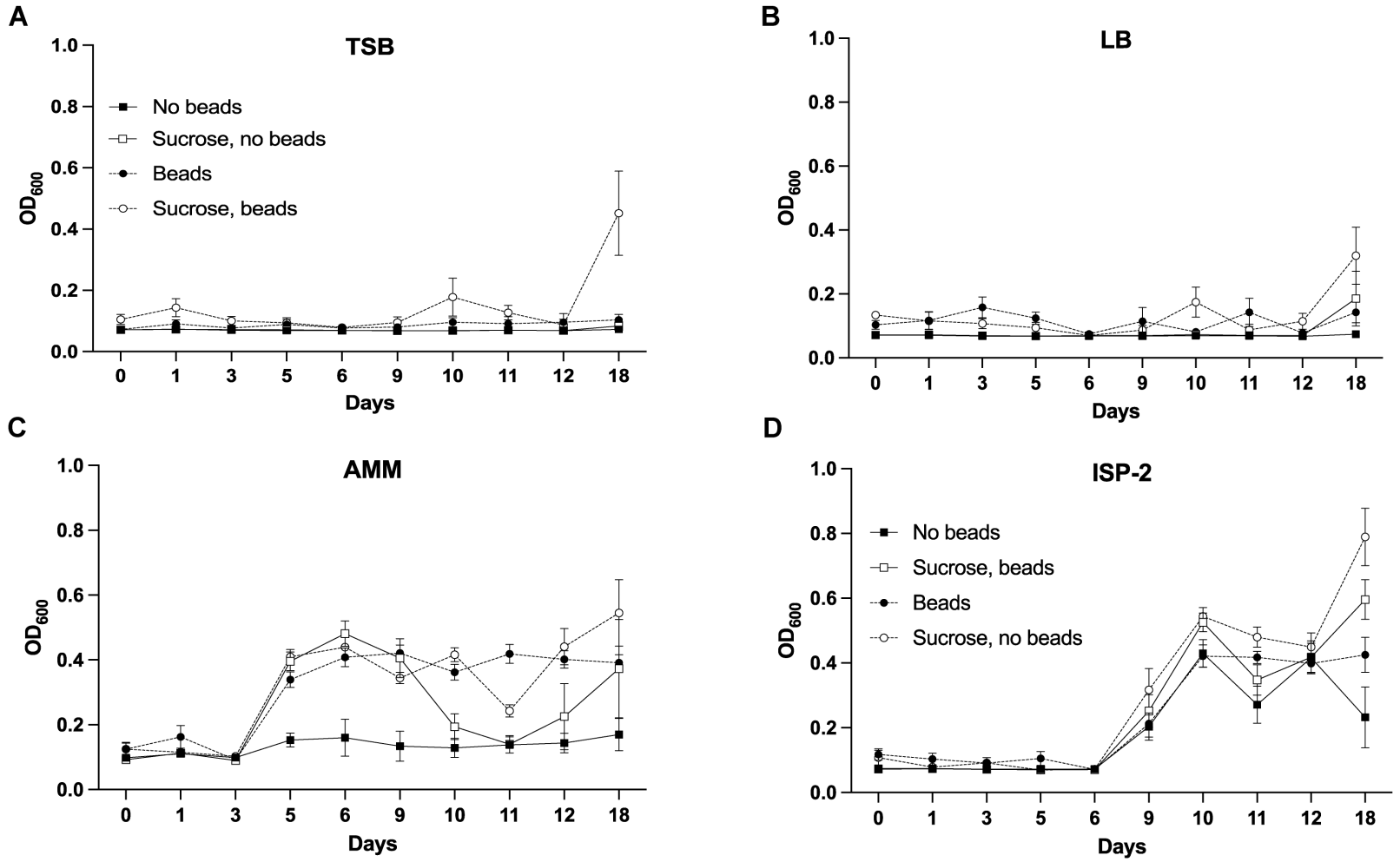

Figure 1: Optimization of growth parameters for *Verrucosispora* sp. CWR15. The isolate was cultured with sterile glass beads (dotted lines, circle) or without (solid lines, square) in four different growth media: **(A)** Tryptic Soy Broth, TSB; **(B)** Luria-Bertani Broth, LB; **(C)** AMM; and **(D)** International *Streptomyces*-2 broth, ISP-2; with 20% sucrose solution (unfilled) or without supplementation (darkened). Data is presented as the average of 5 technical replicates with error bars shown ±SEM.

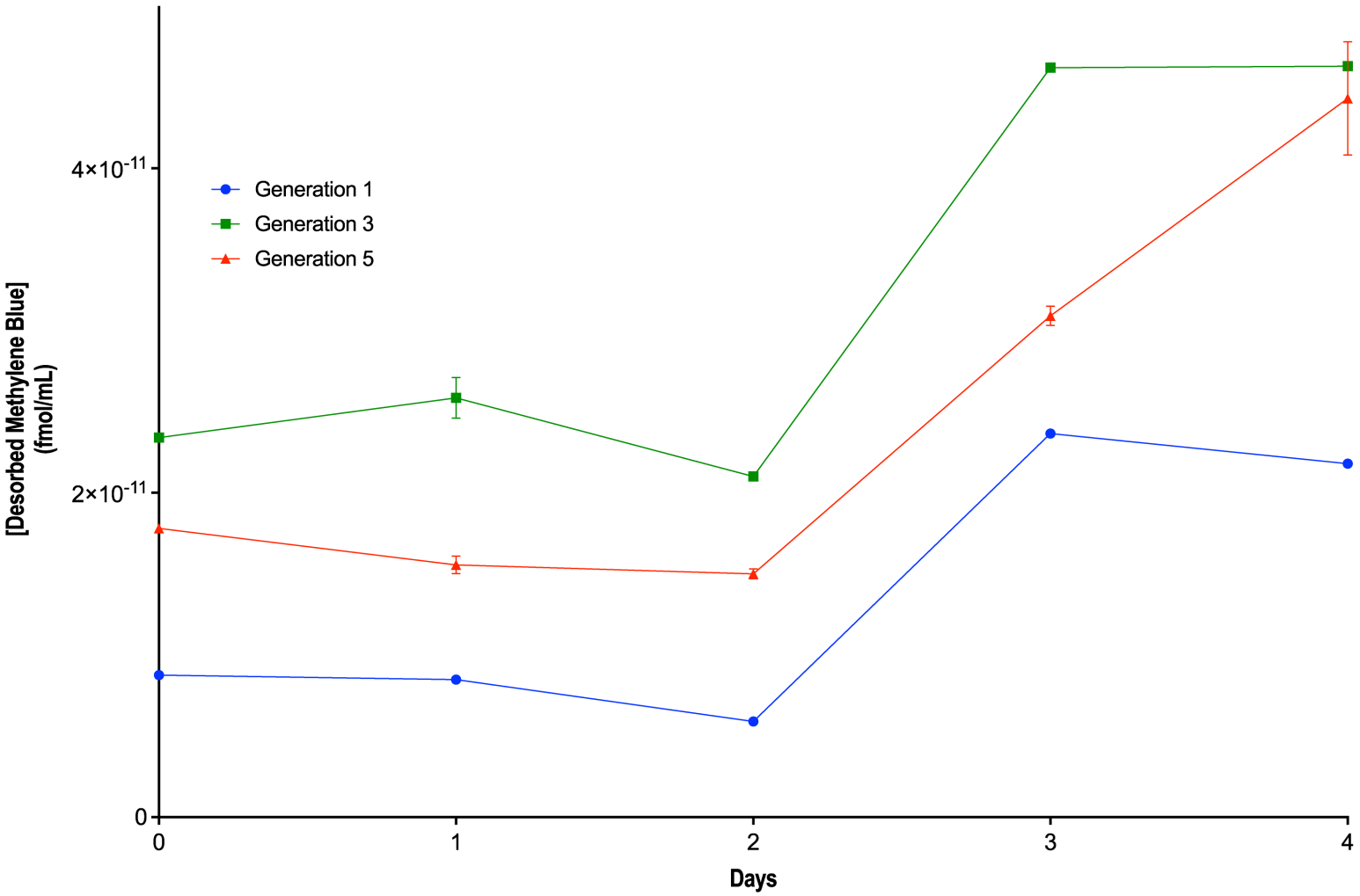

Figure 2: Increased domestication increases biomass. Growth of three generational isolates (1, blue circle; 3, green square; 5, red triangle) was monitored by a modified methylene blue absorbance protocol. Data is presented as the average of 3 technical replicates with error bars shown ±SEM.

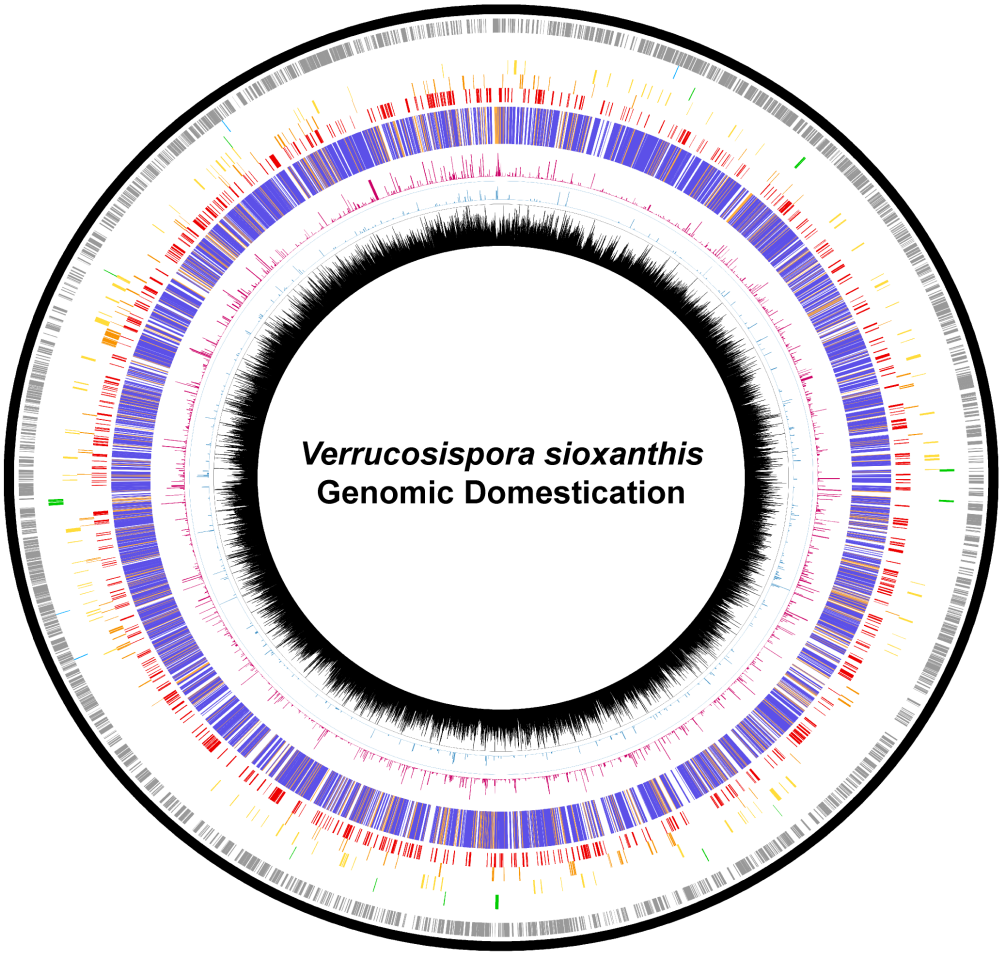

Figure 3: Summary of domestication-induced genomic changes in *Verrucosispora sioxanthis*. Circos plot identifying genomic positions of domestication interest. From outermost ring to innermost: contig-containing karyotype (thick black line), uncategorized KEGG genes (grey), KEGG-designated human diseases/antibiotic resistance genes (blue), KEGG-designated cellular processing genes (green), KEGG-designated environmental information processing genes (yellow), KEGG-designated genetic information processing genes (orange), KEGG-designated metabolism genes (red), pseudogene (purple) or non-pseudogene (yellow-orange) status, indel frequency histogram (magenta, scale 0 – 20%), SNP frequency histogram (blue, scale 0-1%), and coverage histogram (black, scale 0– 100X depth).

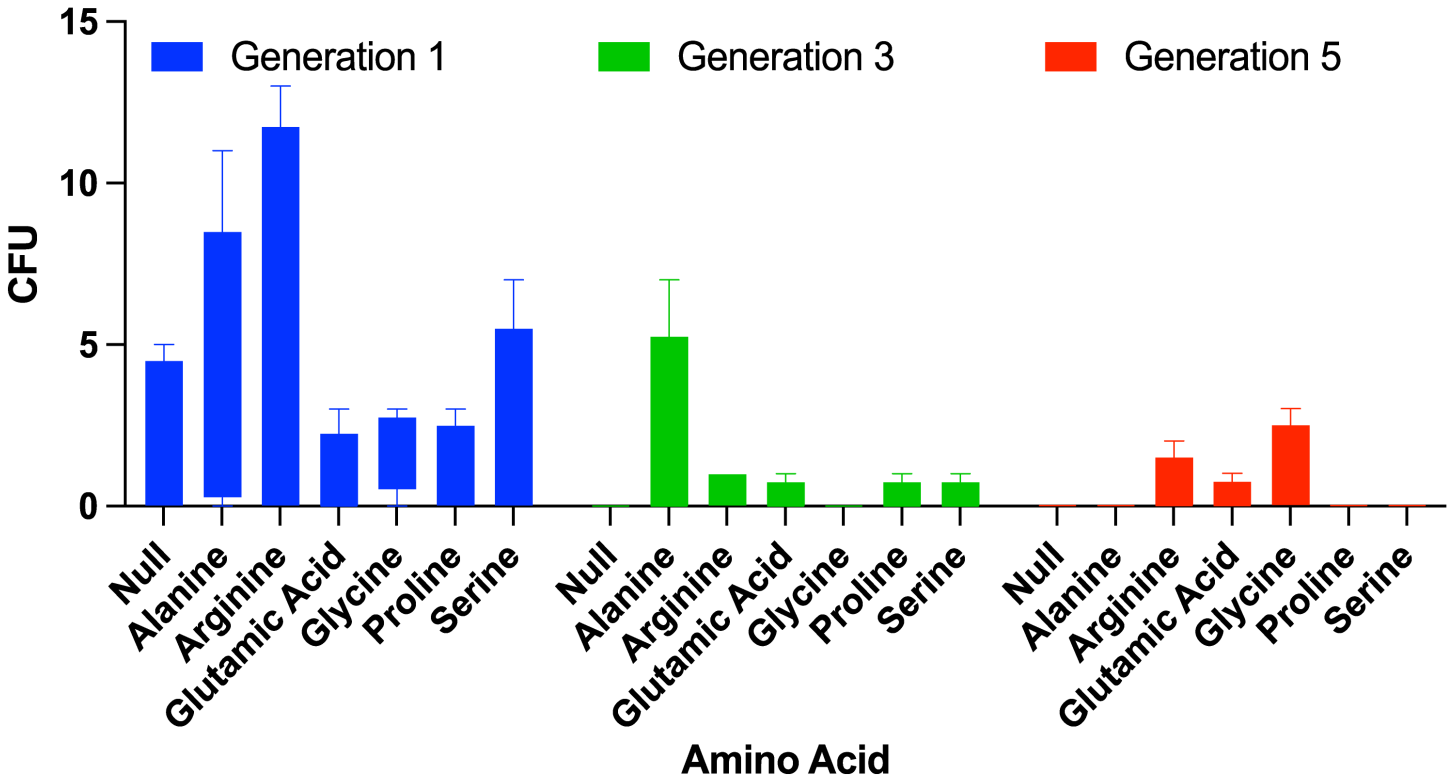

Figure 4: Domestication results in an impaired ability to metabolize diverse amino acids. The three generations were grown in nitrogen-deficient media that was supplemented with the amino acids (0.1% w/v) noted. Cultures were plated in biological duplicate within two technical replicates, and the CFU was enumerated. Data is shown as box and whisker plots with bars denoting the minimum and maximum spread of replicate values (*n*=4).

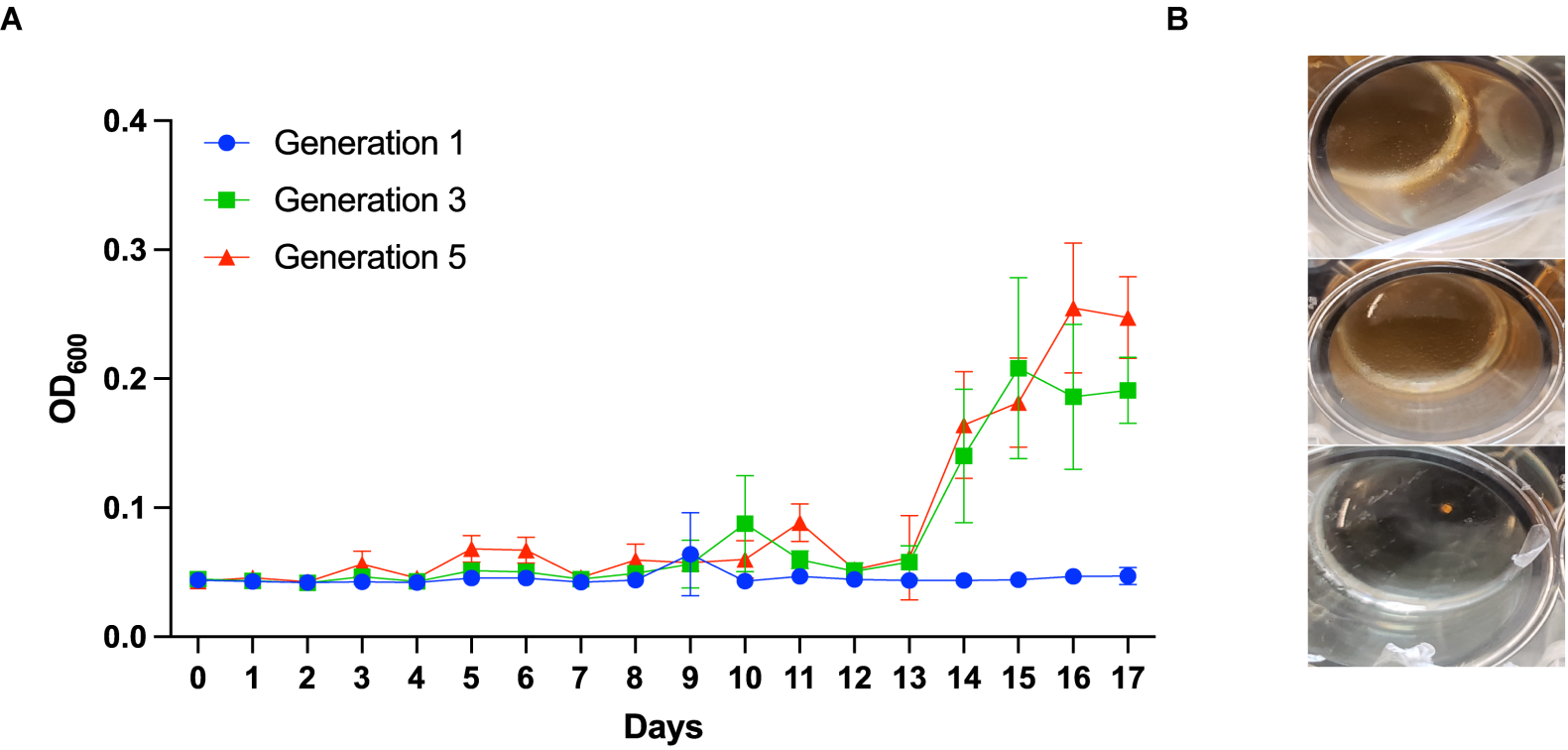

Figure 5: Domestication leads to increased germination rates and growth efficiency. (A) The mean OD_{600} values for each generation are plotted over time as spores initially germinate, and then continue to grow. Each data point is the average of three independent experiments with 6 technical replicates (*n*=18), with 95% CI error bars. **(B)** Endpoint photos of the three generations (1, blue circle; 3, green square; 5, red triangle). No contaminants were observed, and all growth was determined to originate from the heat-treated spore solutions as any contaminating planktonic cells would rapidly overgrow and present as media turbidity.

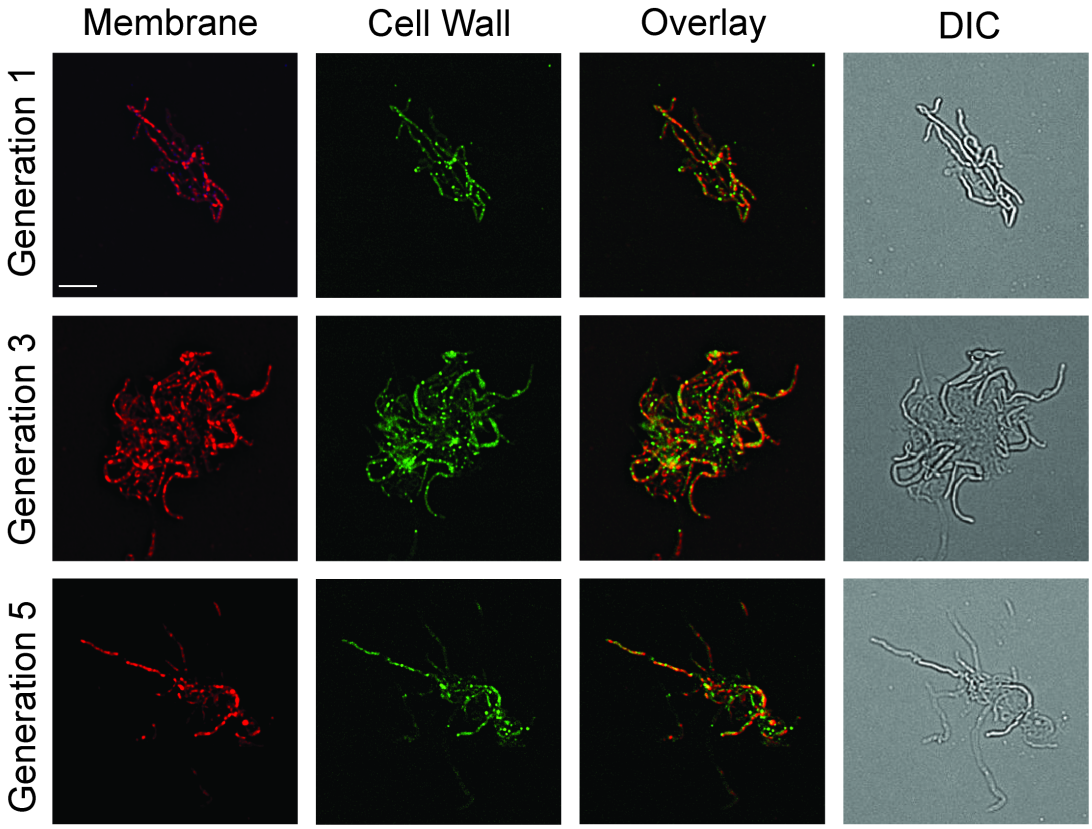

Figure 6: Variable mycelial growth patterns are elicited by *Verrucosispora sioxanthis* domestication. Fluorescence micrographs of cells from Generation 1, 3, and 5 of *Verrucosispora sioxanthis*. Left to right columns show cells with FM4-64 membrane stain (red), BODIPY FL cell wall stain (green), overlay, and DIC. Scale bar is 5µm.

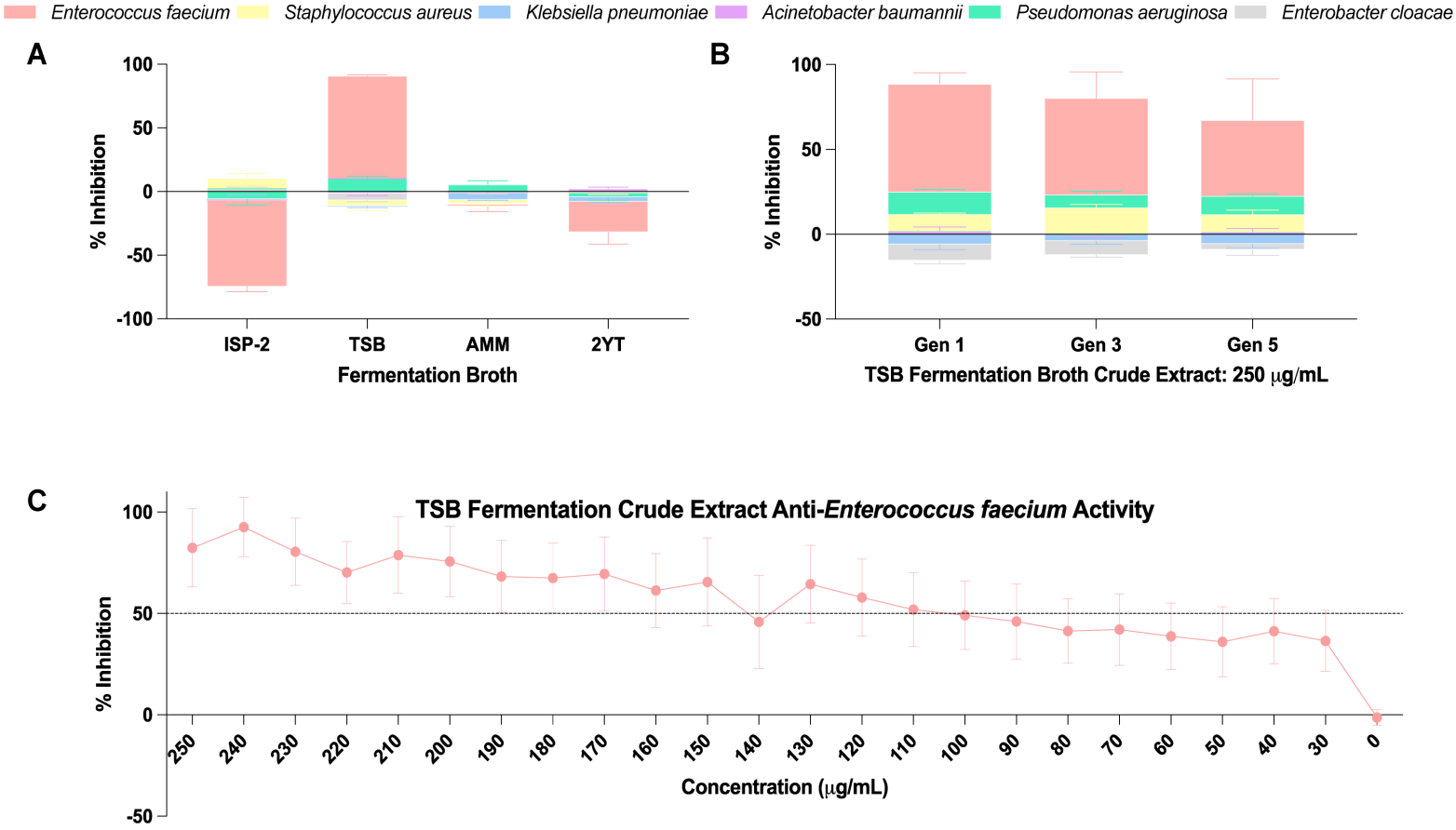

Figure 7: Domestication limits the antibacterial capacity of *Verrucosispora sioxanthis.* G1 was grown for 3 weeks in: International *Streptomyces* Project 2, ISP-2; Tryptic Soy Broth, TSB; AMM; or 2X Yeast Extract Tryptone, 2YT. Crude extracts were generated and assayed against a panel of ESKAPE pathogens for antibacterial activity at a concentration of 250 µg/mL **(A)**. As in **A**, however all three generational strains were used only in TSB **(B)**. Crude extracts of G1 grown in TSB were tested for dose-dependence against *Enterococcus faecium* **(C)**. Percent inhibition was determined by comparison to DMSO-treated controls. Data is from three biological and two technical replicate experiments. Error bars are shown ± SEM.

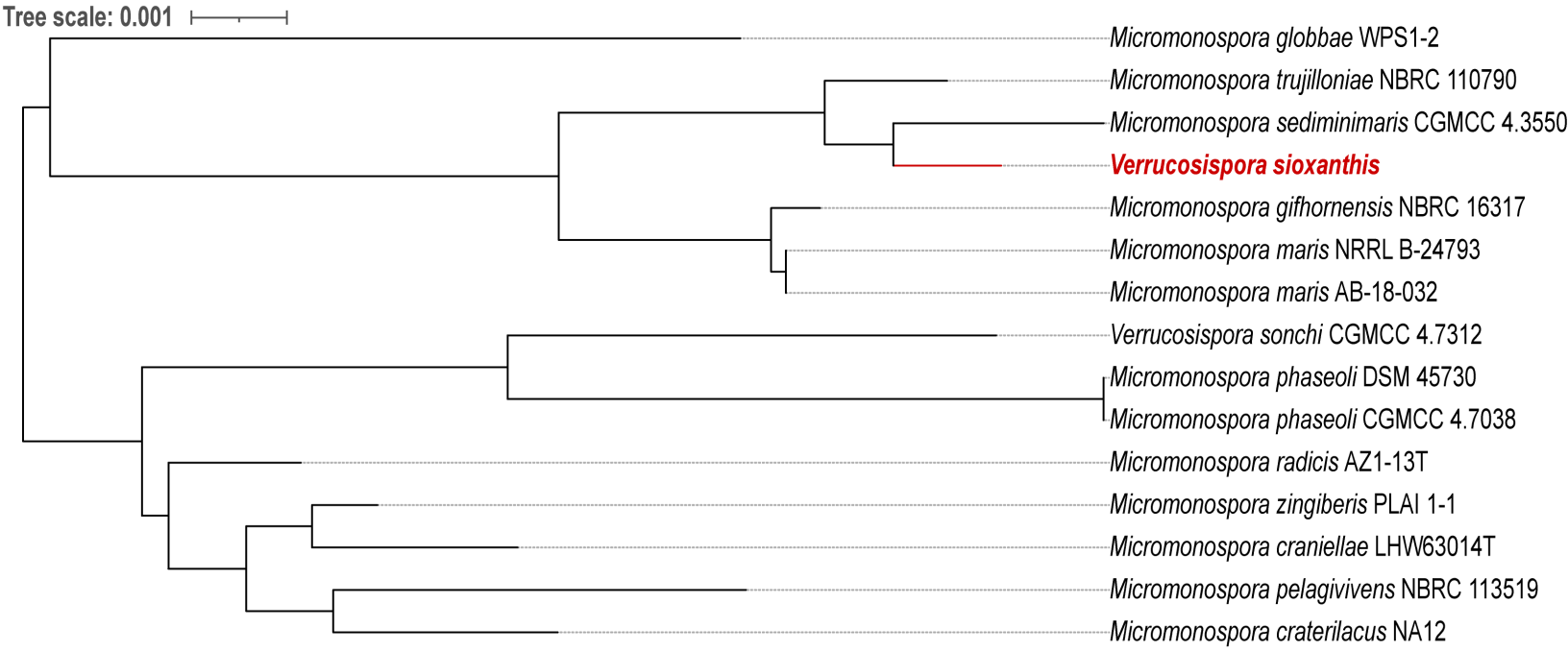

Figure 8: Taxonomy of *Verrucosispora sioxanthis*. Tree inferred with FastME 2.1.6.1 from GBDP distances calculated from 16S rDNA gene sequences. The branch lengths are scaled in terms of GBDP distance formula d5. The numbers above branches are GBDP pseudo-bootstrap support values > 60% from 100 replications, with an average branch support of 66.9%. The tree was rooted at the midpoint. Tree was resolved and edited with Interactive Tree Of Life (iTOL).