Universidad de Oviedo

Instituto Universitario de Oncología del Principado de Asturias

Máster en Biomedicina y Oncología Molecular

Development of modular integrative vectors with SEVA architecture for the diversification of flavonoid biosynthetic pathways in Streptomyces albus

Patrick Laureano McAlpine Álvarez

Faculty of Medicine, Department of Functional Biology, Microbiology Area, BIONUC Research Group

July 2022

Trabajo Fin de Máster

Index

Contents

1.Summary

Flavonoids are a very large and diverse group of polyphenolic compounds that are ubiquitous throughout the plant kingdom and in the human diet through the fruits and vegetables that we eat. Over 6,000 flavonoids have been identified and due to their potent antioxidant, anti-inflammatory, antibiotic, and antitumoral properties, they are considered to have great potential in disease treatment and prevention. Unfortunately, they are produced in exceptionally small quantities by their native plant producers, which makes the agricultural production of these compounds prohibitively expensive. Their chemical synthesis is not much cheaper. Due to the structural complexity and diversity of flavonoids, as well as their stereospecificity, synthesizing them is extremely difficult. For these reasons, biosynthetic production methods in genetically modified microbial cell factories appears to be the most effective and efficient method for their large-scale production.

While other groups have engineered *Escherichia coli* and *Saccharomyces cerevisiae* to produce flavonoids, both organisms require extensive feeding with precursors. Our group has previously demonstrated the efficacy of using *Streptomyces albus* to synthesize a wide variety of flavonoids *de novo*. Still, production yields remain low and diversifying the variety of flavonoids that can be produced remains challenging. The main methods of transforming new biosynthetic gene clusters (BGCs) and genes coding for individual enzymes into *S. albus* consist of either using replicative plasmids or integrative plasmids. Replicative plasmids contain active origins of replication that ensure the plasmids are maintained at high copy numbers. These replicative plasmids are easy to use and yield high levels of transcription, but they are very unstable and can easily be lost if selective pressures, such as antibiotics, are not applied constantly. This makes it difficult to transform the same strain of bacteria with multiple types of plasmids and to grow multiple strains of bacteria in co-cultures, as each plasmid requires a different selective pressure, and each strain must be resistant to all the selective pressures in the culture. Integrative plasmids are plasmids that contain a gene coding for a viral integrase as well as an integration sequence, which is the site at which the plasmid can be integrated into a host genome. These plasmids use the integrase to integrate stably into

the host genome with high locational specificity. Once the plasmid has integrated into the host genome, selective pressures no longer need to be applied. This allows the strain to be further modified in the future without needing to accumulate selective pressures. Also, these bacteria can be used in co-cultures in which no antibiotics are necessary. This greatly increases the potential flexibility of the co-culture, as well as significantly reducing their costs.

Furthermore, integrating the same BGC in multiple locations can increase total flavonoid production by increasing production of the cluster's encoded enzymes. Also, mixing multiple different clusters in the same strain can allow for production of more diverse and complex molecules. Because each integrase only integrates in one or few specific locations in the host genome, expanding the toolbox of potential integrases, and thus the number of sites at which genes can be integrated, is critical for the biosynthesis of all kinds of complex molecules.

In this project we assembled an integrative plasmid using the pSAM2 integrative factor. The use of this integrase was previously established in *S. albus* by other groups but had not been established for use in our group until now. Once assembled, we used it to integrate the naringenin BGC into two different *S. albus* strains. The integration of the plasmid was confirmed using a PCR amplification technique. Once integration was confirmed, the transformed strains were cultured and naringenin production was quantified by high potential liquid chromatography. We found that when the naringenin BGC is integrated into different integration sites, varying amounts of naringenin are produced. We also found that increasing the number of naringenin BGCs in a strain that has been optimized for flavonoid biosynthesis increases naringenin yields.

The establishment of this integration tool will aid in achieving increased production yields of specific flavonoids, as well as the further modification of microbial cell factories to improve the variety and complexity of the flavonoids that can be synthesized. This represents a major step in the establishment of microbial cell factories that can produce flavonoids at industrially viable scales

2. Introduction

2.1 Flavonoids

2.1.1 Functions in Plant and Animal Physiology

Flavonoids have a tremendous variety of functions within the plants that produce them, ranging from protection from ultraviolet radiation, free radicals, bacterial and fungal pathogen, and even regulation of symbiotic interactions (1). It is precisely these functions in plants that makes these chemical compounds so interesting from a biomedical perspective. In animals, flavonoids have a tremendous number of health promoting properties. They are potent antibiotics, anti-inflammatories, antitumorals, and antioxidants (2). Previous research has demonstrated that diets high in flavonoids are associated with reduced risk of cardiovascular disease (3) colorectal cancer (4), and even neurodegenerative diseases (5). Beyond just dietary ingestion of flavonoids, their use as nutraceuticals for cancer treatment and prevention has also been explored and demonstrated extensively in animal models (6–9). Our own BIONUC research group has even demonstrated that including flavonoids in processed meat products reduces their carcinogenicity (10).

As remarkable as these flavonoids are as potential treatments and preventative factors in chronic diseases, their potent function as antibiotics may be of even greater immediate importance. It is well established that antibiotic resistant bacteria are continuing to become more common and are becoming a major threat to human health (11). In the coming years, the race to identify novel antibiotic drug candidates will become increasingly frantic. On top of this, the importance of microbial ecosystems (microbiomes) to human health is also becoming increasingly clear (12), and the negative impact of antibiotic usage on our microbiome has already been demonstrated to have detrimental effects on human health (13). Flavonoids offer a potential tool for ameliorating both problems. They have been shown to not only be effective antibiotics against resistant strains, but also to inhibit virulence factors and biofilm production, as well as reverse

antibiotic resistance when paired with other antibiotic drugs(14). They also exhibit positive effects on the human intestinal microbiome. They recompose the microbiome by reducing growth of pathogenic bacteria and increasing growth of commensal bacteria (15).

For all the stated reasons, flavonoids have the potential to become critical nutraceutical or even pharmaceutical tools in the treatment and prevention of disease. For now, the limiting factor preventing their use is the difficulty in producing flavonoids in sufficient quantities. This is one of the areas of investigation in which BIONUC is taking part, and the goals of this *Trabajo Fin de Máster* tackle this problem directly.

2.1.2 Limitations in Flavonoid Production

Currently, there are three methods for flavonoid production, extraction from plants, chemical synthesis, and heterologous biosynthesis in microbes. While flavonoids are produced naturally in plants, this production occurs in very small quantities (in the μmol g^{-1} of plant material range) (16). Because of this, agricultural production and extraction of flavonoids is extremely expensive and land intensive. Chemical synthesis is proving extremely complicated as well due to the stereospecificity and the unique methylation or hydroxylation patterns that are characteristic of different flavonoids. While it is easy to methylate every carbon on the flavonoid backbone using chemical techniques, methylating only specific carbons is extremely expensive and laborious (17). For these reasons, biosynthesis in microbial cell factories is proving to be the most effective method for flavonoid production. Past work by the BIONUC group has demonstrated *de novo* production of flavonoids in the 30-50 μg l^{-1} range in liquid cultures of *Streptomyces albus* (18). Moreover, our group has shown the tremendous range of flavonoids that can be produced using biosynthetic methods, having already produced naringenin, garbanzol, kaempferol, fustin, quercetin, apigenin, luteolin, and eriodyctiol, among others (18–20). Our group has worked tirelessly to increase flavonoid production over the last few years and data that will be published soon will demonstrate the remarkable improvements in flavonoid yields. Even so, greater yields are necessary if flavonoids are to become viable sources of nutraceuticals and even pharmaceuticals.

2.1.3 Structure and Classification

Flavonoids are a highly diverse group of compounds that are ubiquitous throughout the plant kingdom. More than 6,000 have been identified and they are characterized by their highly conserved structure. In general, they consist of 2 polyphenolic benzine rings (rings A and B) joined by a heterocyclic pyran ring (ring C) (21). Flavonoids can be divided into subgroups depending on the location and stereospecificity of the hydroxyl groups on the flavonoid backbone. Six of the most common subgroups consist of flavones, flavanols, anthocyanidins, flavanones, chalcones, and isoflavones figure 1.

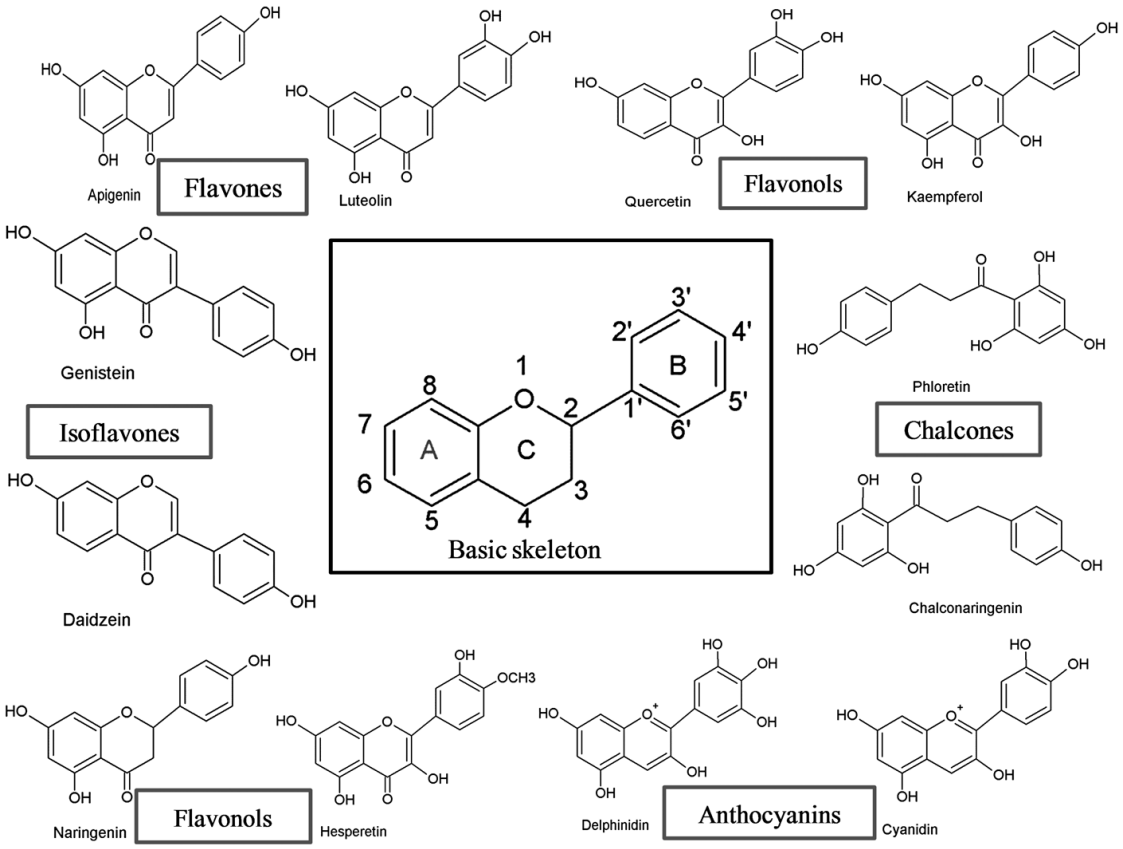


Figure 1. Six subclasses of flavonoids are shown, with the hydroxyl and carbonyl groups that are characteristic of each. Letters A and B represent the two benzene rings that are joined by the pyran ring C. Adapted from Panche and Diwan, 2016.

2.1.4 Natural Synthesis in Plants

Flavonoids are naturally produced primarily in plants as secondary metabolites. Although there is a tremendous variety of flavonoids and of metabolic pathways that produce them, the precursors for all flavonoids are synthesized via the acetate pathway and the shikimic acid pathway (22). In the shikimic acid pathway phosphoenol-pyruvic acid and erythrose-4-phosphate react to form shikimic acid, which is a precursor for aromatic amino acids and phenyl propanoids. These then react with acetyl coenzyme A to form chorismate, which is further metabolized to form tyrosine and other aromatic amino acids such as phenylalanine. Although multiple flavonoid biosynthesis pathways exist starting from various aromatic amino acids, they generally follow a similar path. For example, from tyrosine, four plant enzymes catalyze its conversion to naringenin, which is one of the central flavonoids from which others are derived. First, tyrosine ammonia lyase converts tyrosine into coumarate, then 4-coumaric-acid CoA ligase converts coumarate to coumaroyl-CoA, then chalcone synthase converts coumaroyl-CoA to naringenin chalcone using 3 molecules of malonyl-CoA, and finally chalcone isomerase converts naringenin chalcone to naringenin Figure 2 (23).

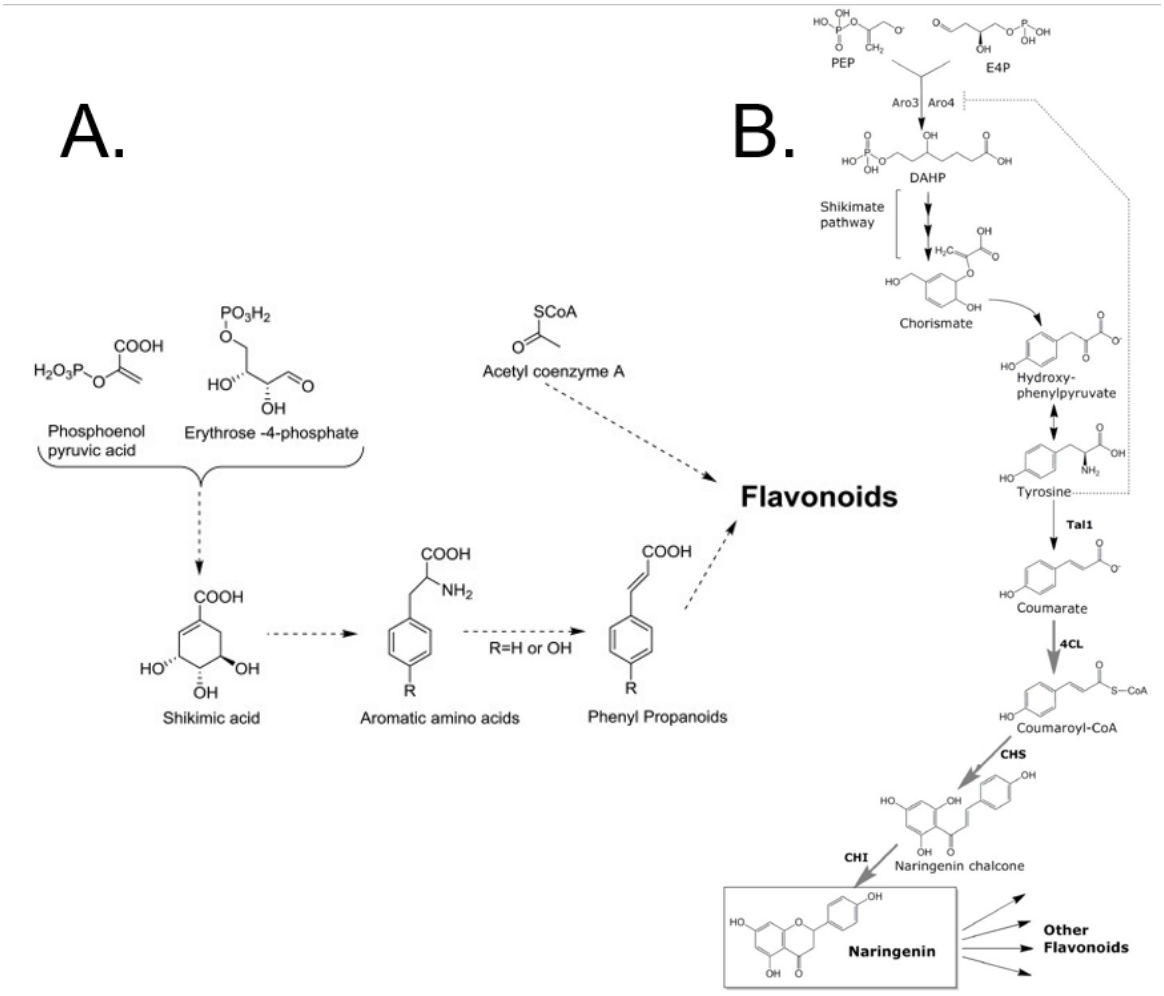


Figure 2. A: Shikimic acid pathway of flavonoid precursor biosynthesis, adapted from Nabavi and Šamec, 2020. B: Naringenin biosynthesis pathway via tyrosine, TAL1 is tyrosine ammonia lyase, 4CL is 4-coumaric acid-CoA ligase, CHS is chalcone synthase, and CHI is chalcone isomerase. Adapted from Koopman and Beekwilder, 2012.

2.1.5 Naringenin

Naringenin is a flavanone that is particularly common in citrus fruits. Like other flavonoids, it has been shown to have many health promoting properties (24), but what makes it of particular interest for flavonoid biosynthesis is its role as a central node in flavonoid metabolism. As already described, tyrosine can be easily converted to naringenin, and once naringenin is made, multiple other flavonoids can be further derived from it. Apigenin, luteolin, aromadendrin, taxifolin, kaempferol, quercetin, and myricetin can all be derived from naringenin within 3 enzymatic steps Figure 3 (19). For this reason,

naringenin production is a critical bottleneck in the production of other diverse flavonoids. Until naringenin yields can be improved, other flavonoids will not be producible at large scales, and thus a primary focus of flavonoid biosynthesis revolves around increasing naringenin production.

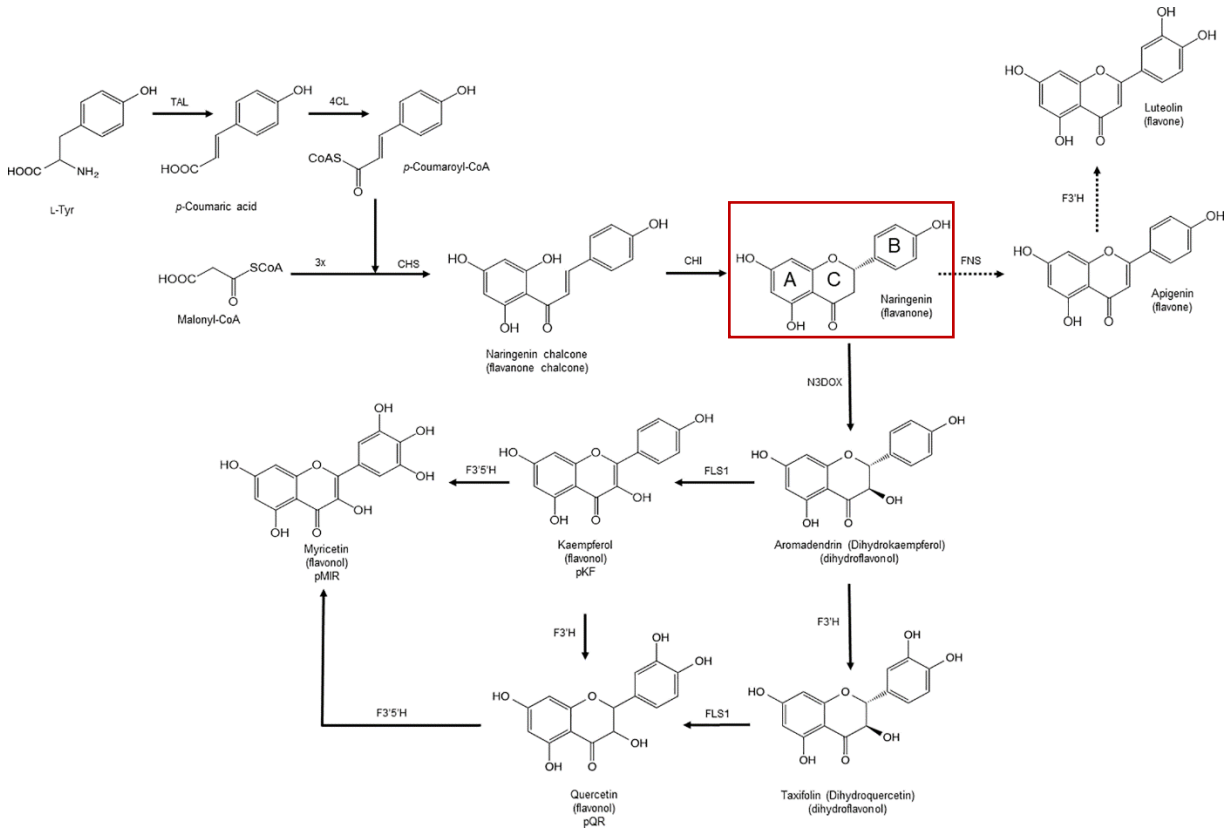


Figure 3. Flavonoid diversification from naringenin as a central node. Adapted from Marín and Gutiérrez-del-Río 2018.

Furthermore, the metabolic pathway from tyrosine to naringenin is well understood, the enzymes involved have been characterized, and the genes coding for these enzymes have been sequenced. This makes biosynthetic production of naringenin highly feasible. Still, achieving high yields remains difficult, which severely limits production of further flavonoids. The first step in achieving industrially relevant flavonoid yields is producing large amounts of naringenin.

2.2 Heterologous Biosynthesis

2.2.1 *Streptomyces albus*

Streptomyces albus is a gram-positive bacteria and part of the *Actinomycetales* order (phylum *Actinobacteria*). The strain that we use for flavonoid biosynthesis in BIONUC is *S. albus* J1074. Its genome contains just under 6,000 genes, making it a relatively simple organism to modify, when compared with the plants that normally produce flavonoids. Also, it grows relatively quickly in liquid culture compared with other actinomycetes and it is very receptive to foreign DNA, making it very easy to transform (25). While *S. albus* may be easy to work with compared with other actinomycetes, it is still not nearly as usable as other model microbes such as the bacteria *E. coli* and the yeast *S. cerevisiae*. What makes the actinomycetes attractive hosts for flavonoid biosynthesis is their incredible metabolic capabilities. Actinomycetes produce up to two thirds of all the industrially relevant bioactive molecules used today. They also have extensive protein post translational modification mechanisms that allow for the effective use of enzymes from eukaryotic origins (26). They already contain much of the enzymatic requirements needed to metabolize flavonoid precursors and as such are the ideal microorganisms to produce flavonoids *de novo.*

Although *S. albus* is an ideal host for flavonoid biosynthesis it is not without its downsides. Its genome contains a remarkably high ratio of guanine-cytosine nucleotides (73.3%) (25) which can complicate gene synthesis. Also, it being Gram-positive complicates the transformation process since the thick cell wall must be circumvented. Even so, both problems are surmountable. The use of bioinformatic tools to successfully optimize genes for *S. albus* expression can also be used to reduce their G-C content and facilitate their synthesis. Also, several strategies have been developed for *S. albus* transformation such as using intermediary *E. coli* to conjugate the synthetic genes into *S. albus,* as well as the enzymatic digestion of the cell wall to facilitate DNA uptake.

2.2.2 Transformation and Biosynthesis

There are generally three methods for transforming *S. albus* into flavonoid producing factories, CRISPR-Cas9 genome modifications, replicative plasmids, and integrative plasmids. CRISPR technologies offer the ability to cut the bacterial genome with remarkable specificity and flexibility but getting foreign DNA to recombine in the cut location is extremely challenging. For this reason, CRISPR technologies are largely used to knock out gene clusters that code for enzymes that compete with the biosynthetic pathway. Using CRISPR to knock in desired genes is not practical and thus gene insertion is usually conducted using one of two plasmid methods.

Replicative plasmids are particularly useful due to the ease with which they can transform the host organism. These plasmids typically contain the biosynthetic gene cluster to be inserted into the host, a high copy number origin of replication, and a gene coding for resistance to a specific selective pressure such as an antibiotic. Because of the large number of copies that these plasmids maintain within the host cell, they are very effective in transforming and yield many desired gene cluster transcripts within the host cell. Although high copy number replicative plasmids offer ease of transformation and high expression levels, they do have their downsides. Like any plasmid, they act as a negative resource expenditure on the host cell and thus can be expelled from the cell if selective pressures are not constantly applied. While this may seem trivial, there are multiple reasons why working without selective pressures is desirable, including environmental and economic ones when targeting large scale production. Even resistant bacteria can have their growth rate in culture reduced in the presence of antibiotics. Furthermore, the desired biosynthesis product may require the insertion of multiple gene clusters, which increases the number of plasmids that must be contained within the host cell. This forces the use of multiple antibiotics at the same time to select for all the plasmids simultaneously. Also, the use of bacterial consortia to increase product yields is currently in development, and if antibiotics are to be used then all the members of the consortium must be resistant (27). Thus, having to use antibiotics or other selective pressures greatly complicates the ability to work with complex biosynthetic systems. Furthermore, the purpose of generating flavonoids through biosynthetic methods is to

reduce the cost of production. Antibiotics are very expensive and are generally not desirable for use in industrial settings.

A third option for transforming *S. albus* is the use of integrative plasmids. Unlike replicative plasmids, these plasmids contain a gene coding for an integrase which, when translated, produces an enzyme that integrates the plasmid into a specific location in the host genome. Because the plasmids are fully integrated into the host genome, there is no risk of the genes being lost in the absence of selective pressures. This improves the long-term viability of the generated strain and makes co-culture with other microorganisms more viable. Furthermore, each future transformation only needs to be done with one selection mechanism since the previous transformations cannot be lost.

A major downside to the use of integrative plasmids is that each integrase can only act upon a very specific DNA sequence. For this reason, integrative plasmids can only integrate into one or very few locations within a single host. Thus, to integrate multiple gene clusters or conduct multiple transformations, each new plasmid must contain a gene for a different integrase. Also, transcription of the gene products from only one integration can be much lower than from high copy number replicative plasmids. Integrating the same gene cluster into multiple sites in the genome increases translation of its encoded enzymes and could increase production of the desired flavonoid. Furthermore, integration of multiple different gene clusters into a single host can allow for greater diversification in the types of flavonoids that can be made. Thus, in order to perform complex modifications upon the host strain, it is necessary to have multiple integrases established and optimized, so adding new integrases to the transformation tool kit is of vital importance (28).

2.2.3 Integrases

Integrases are enzymes that integrate two pieces of DNA at specific locations. They typically recognize an *attB* site that represents the location on the host DNA at which the foreign DNA will be inserted, and an *attP* site on the foreign DNA that represents the location at which the foreign DNA being inserted will interact with the receiving host DNA. Generally this means that the *attB* site is the location in the host genome in which the

plasmid will be inserted and the *attP* site is the location on the plasmid at which it will be split and have its ends joined into the host genome Figure 4(29).

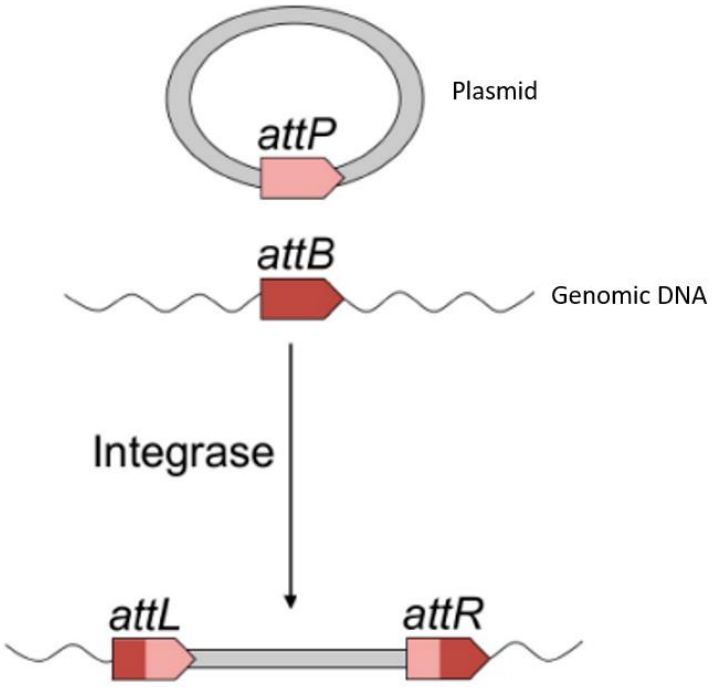


Figure 4. Integrase function in integrating a plasmid into a genome at the *attP* and *attB* sites. Adapted from Merrick and Zhao, 2018.

The majority of integrases come from viral origins although some have unknown origins (30). While many integrases have been used for *Streptomyces* genetic engineering, only the integrases φC31 and φBT1 have been used extensively by our group. In this work, we develop a platform for usage of the pSAM2 integrase within our group. pSAM2 was first discovered in *Streptomyces ambofaciens* as an integrated conjugative plasmid by Boccard and Smokvina in 1989 (31). Since then, it has been used for genetic engineering in *S. albus* and its single *attB* integration site in the *S. albus* genome has been documented (32).

2.2.3.1 Integrase Applications and Questions that Remain

As stated previously, integrases only integrate into one or few locations in the host genome. Thus, although integration offers the advantage of strain stability, it is difficult to get high yields of the desired product because the BGCs exists in very small numbers of copies. One technique to increase yield is to integrate the same BGC into multiple locations in the genome to increase its expression (28). While this may increase production to an extent, it is likely to only work up to a limited capacity since gene expression cannot circumvent other bottlenecks such as limited nutrient and precursor quantities.

A consideration when using integrases is that different areas of the bacterial genome have different levels of transcriptional activity. Others have previously demonstrated this phenomenon in *E. coli* (33). As such, different integrases may yield different amounts of transcriptional activity, which could significantly alter product yields. Selecting specific integrases that integrate into the most active sites of the host genome could be an easy way to increase yields, while optimizing metabolic pathways such that the genes coding for precursor biosynthesis are integrated into the most active sites and the genes coding for final product biosynthesis are integrated into less active sites may be an effective and rational way to prevent production bottlenecks and force the reaction dynamics toward the desired product.

3. Objectives

The objectives of this *Trabajo Fin de Máster* are three-fold. The first objective was to establish a modular, easily modifiable plasmid in the SEVA architectural model that contained hygromycin resistance genes for selection and the gene for pSAM2 integrase for integration. Up to this point the pSAM2 and hygromycin resistance genes have not been in use in the BIONUC group, so developing this tool will allow for the easy integration of biosynthetic gene clusters (BGCs) into a novel location in the *S. albus* genome. This will allow for greater diversification in the metabolic pathways that can be generated and

thus increase the variety of flavonoids that can be produced through *S. albus* biosynthesis.

The second objective was to document differences in flavonoid yield when the BGC is integrated into different areas of the genome using different integrases. As stated previously, we hypothesized that varying levels of transcriptional activity in different regions of the genome could cause differences in flavonoid production. These differences in flavonoid production could be used to optimize production of complex flavonoids by ensuring that the BGCs coding for precursor production are integrated into the most active areas of the genome. This would prevent bottlenecks downstream in the metabolic pathway.

The third objective was to demonstrate that integrating the same BGC multiple times into one strain could be an effective method for increasing flavonoid biosynthesis. This method has been demonstrated in other models (28), but has not yet, to the best of our knowledge, been used for flavonoid biosynthesis. Of particular interest was to see at which point this method failed to provide a further increase in flavonoid product, indicating that a separate bottleneck had been reached.

4. Materials and Methods

4.1 DNA manipulation

Gene Optimization and Synthesis

The flavonoid BGCs are native to plants and thus need to be optimized both for use in *S. albus* and for our DNA manipulation methods. The hygromycin resistance gene is native to *Streptomyces hygroscopicus* (34) which is similar enough to *S. albus* that the gene only needed to be optimized for our DNA manipulation methods. First, the IDT organism codon optimization tool was used to modify the codons in such a way that availability of tRNAs in *S. albus* were maximized without altering the amino acid sequence. Next, nucleotides were changed to eliminate recognition sites for *BbsI* and

*Bsa*I restriction enzymes to ensure that the genes would not be cut in undesired locations during Molecular Cloning. Next, the IDT gene synthesis tool was used to evaluate the viability of synthesizing the desired genes. Using the tool, nucleotides were changed to eliminate hairpins, long repetitive sequences, and to reduce the guanine cytosine content of the genes. Once the genes were confirmed to be synthesizable, the sequences were sent to Explora Biotech (Italy) to be synthesized.

Plasmid DNA Extraction from *Escherichia coli* (Mini Prep)

All mini preps from *E. coli* were conducted as follows. 1.5 mL of culture were collected in 1.5mL Eppendorf tubes. The tubes were centrifuged at 10,000 rpm for 1 minute on a benchtop centrifuge and then the supernatant was removed. The remaining pellet was resuspended in 100 µL of solution I (glucose 1M, EDTA 0.5 M, Tris-HCl 1M, pH 8). 200 uL of solution II (NaOH 1 M, SDS 10%) was then added to lyse the *E. coli* and the tubes were incubated at room temperature for 5 min. To precipitate the SDS, cellular debris, and genomic DNA, 150µL of solution III (potassium acetate 3 M, acetic acid 11.5%) was added to the tubes. They were then mixed and centrifuged at 14,000 rpm for 10 min. The resulting supernatant was then transferred to a new Eppendorf tube and 400 uL of isopropanol was added. The tubes were then centrifuged again for 10 min at 14,000 rpm and the resulting supernatant was removed. 100 uL of 70% ethanol was added and the tubes were then centrifuged for 1 min at 14,000 rpm. The resulting supernatant was removed, and the tubes were incubated at 37 ^{o}C until the DNA pellet was completely dry. The pellet was then resuspended in 50 uL of TE buffer (Tris-HCl 10 mM, EDTA 1 mM, pH 8) and quantified using a Qubit (Thermo Fisher Scientific, USA) (35).

***S. albus* genomic DNA Extraction**

DNA was released from *S. albus* spores by incubating 1 µL of spores in 45 µL of 50 mM NaOH at 95 ^{o}C for 10 min. Following 10 min the NaOH was neutralized by addition of 5 µL of 1 M Tris-HCl pH 8.

Polymerase Chain Reaction (PCR)

PCR reactions to amplify fragments from plasmids for further cloning were conducted using the Herculase II Fusion DNA Polymerase (Agilent, USA) and a Veriti 96 well thermocycler (Applied Biosystems, USA). In each reaction, 10 µL of 5x Herculase II reaction buffer, 0.5 µL of dNTP mix, 0.5 µL of DNA template, 1.25 µL of each primer at a concentration of 10 µM, 1 µL of Herculase II fusion DNA polymerase, 2.5 µL of dimethyl sulfoxide (DMSO), and distilled water to 50 µL were mixed. The cycling conditions were set to 1 cycle at 95 ^{o}C for 2 min, 30 cycles of 95 ^{o}C for 20 s (denature), optimized primer annealing temperature for 20 s (annealing), and 72 ^{o}C for 30 s per kb of DNA (extension). Finally, the reaction ended with 1 cycle at 72 ^{o}C for 3 min.

To conduct PCRs on *S. albus* genomic DNA we use the Terra Polymerase (Takara, USA) which is capable of amplifying DNA specifically in a wide range of environments. For this reason, the DNA does not need to be pure, it only needs to be released from the cells. Thus, no further treatment of the DNA was necessary post cell lysis in NaOH. 12.5 µL of Terra buffer, 0.75 µL of each primer at a concentration of 10 µM, 5 µL of DNA, 0.3 µL of Terra polymerase, and ultra-pure water up to a final volume of 25 µL was mixed. The mixture was then placed in the thermocycler at 98 ^{o}C for 10 s followed by 40 cycles of 98 ^{o}C for 10 s and 68 ^{o}C for 1 min/kb of amplicon product.

DNA cloning

To form the initial pSAM2int hyg C15D plasmid, we used a protocol adapted from Gibson et al. known commonly as the Gibson Assembly (36). Using the above PCR protocol, we amplified the desired regions from various already assembled plasmids. The Herculase II fusion DNA polymerase creates 5’ overhangs which can be used to join homologous ends. Because our plasmids are made in the SEVA architecture, the 5’ overhangs are made in the UNS sites so that the different regions of the plasmids can easily be joined.

Once the desired regions had been amplified, the PCR amplicons were purified by gel electrophoresis. The desired bands were excised from the gel, purified using the GeneJet Gel kit (Thermo Fisher Scientific), and quantified using a Qubit DNA quantifier (Thermo Fisher Scientific). The New England Biolabs Gibson Assembly calculator was then used to calculate the amount of each amplicon to mix in the final reaction, and the

New England biolabs Gibson Assembly kit was used to join the amplicons and form the final plasmid with an incubation period of one hour at 50 ^{o}C.

We used the CIDAR modular cloning protocol (37) to remove the LacZ operon from the gadget in the pSAM2int hyg C15D plasmid and insert the four genes coding for the naringenin BGC. The 4 enzyme coding genes were stored in plasmids containing level 1 gadgets that contained the gene coding for the enzyme along with promoters and transcription factor binding sites. These level 1 plasmids were generated previously.

To conduct the CIDAR method, the concentration and size of our level 2 plasmid, along with those of the previously generated level 1 plasmids were entered into a MoClo calculator (in house) to calculate the amount of each plasmid to mix in the final reaction. The plasmids were mixed in the calculated amounts along with 1 µL of *Bbs*I restriction enzyme (Thermo Fisher Scientific, USA),1 µL of T4 ligase (Thermo Fisher Scientific, USA), 4 µL of T4 buffer (Thermo Fisher Scientific, USA), and 4.75 µL of ultra-pure water. The reaction mixture was then incubated in a thermocycler with the following settings: 25 cycles of 37 ^{o}C for 1.5 min and 16 ^{o}C for 3 min, followed by 5 min at 50 ^{o}C and 10 min at 80 ^{o}C.

Restriction Digestion

Restriction digestions were carried out by mixing 4 µL of plasmid with 1 µL of digestion buffer, 0.5 µL of *Sal*I Fast restriction enzyme (Thermo Fisher Scientific, USA), 0.1 µL of RNAse enzyme, and 4.4 µL of ultra-pure water. They were then incubated at 37 ^{o}C for 30 min.

Gel Electrophoresis

Gel electrophoresis was done in a 0.7% agarose gel containing 5% Red Safe Dye (iNtRON Biotechnology, South Korea). PCR products had 6x loading buffer (Thermo Fisher Scientific, USA) added to them prior to electrophoresis while digestion products did not need this because the digestion buffer (Thermo Fisher Scientific, USA) already acts as loading buffer. 5-10 µL of sample were added to the wells in the gel and 5 µL of standard PCRBIO Ladder II (PCR Biosystems, UK) was added to the ladder positions in the gel. Samples were run at 120 V for 30 min.

Sequencing

Plasmids that were sent for sequencing were first purified from *E. coli* using the Gene Jet Mini Prep kit (Thermo Fisher Scientific, USA) for maximum purity. 100 ng/µL of kit-purified plasmid DNA was mixed with 10 pmol/µL of primer. This mixture was then sent to Stab Vida (Lisbon Portugal) for Sanger sequencing. Analysis of reads was done using SnapGene software (Dotmatics, UK).

Plasmids

Plasmid names and information are available in Table 1.

Plasmids		
Name	Function	Origin
VWBint ampi-kana C15D	Source of molecular cloning gadget and origins of replication for pSAM2int hyg C15D assembly	BIONUC, not yet published
pSEVA181 HygR	Source of hygromycin resistance gene for pSAM2int hyg C15D assembly	Explora Biotech (Italy)
pSEVA181 pSAM2	Source of pSAM2 integrase gene for pSAM2int hyg C15D assembly	Explora Biotech (Italy)
pSAM2int hyg C15D	Plasmid that integrates into the pSAM2 integration site, is resistant to hygromycin, and contains a gadget for a four gene biosynthetic gene cluster	This project via Gibson Assembly
pSEVABT1intApra1A12-permE-RiboJRES-TAL-ttsbiB	Level 1 plasmid for assembling the tyrosine ammonia lyase gene into a cluster	BIONUC, not yet published
pSEVABT1intApra2AE3-sf14-RiboJRBS-4CL-ttsbiB	Level 1 plasmid for assembling the 4-coumaric acid-CoA ligase gene into a cluster	BIONUC, not yet published
pSEVABT1intApra3AI4-Sp25-RiboJRBS-CHS-ttsbiB	Level 1 plasmid for assembling the chalcone synthase ligase gene into a cluster	BIONUC, not yet published
pSEVABT1intApra4AIS-Sp43-RIBOJRBS-CHI-ttsbiB	Level 1 plasmid for assembling the chalcone isomerase gene into a cluster	BIONUC, not yet published
pSAM2int hyg NAR	Level 2 plasmid that was used to integrate the naringenin BGC into the *S. albus* genome at the pSAM2 integration site	This project via molecular cloning

Table 1. Features of plasmids used in this work.

Primers

Primer names and corresponding sequences are available in Table 2.

Name	Sequence
For Gibson Assembly Amplicon Generation	
UNS1 rev	GAGACGAGACGAGACAGCCTGAGAATG
CRISPR F1 UNS1 fw bis	CATTACTCGCATCCATTCTCAGGCTGTCTCGTCTCGTCTCCTTTTCCGCTGCATAACCC
CRISPR F1 UNS2 rev	GCTTGGATTCTGCGTTTGTTTCCGTCTACGAACTCCCAGCTCATGGCTCTGCCCTCGG
CRISPR F2 UNS2 fw	GCTGGGAGTTCGTAGACGGAAACAAACGCAGAATCCAAGCGGCCGGCCCTACGCGACCGCTGTGTCGA
UNS3 rev	CGACCTTGATGTTTCCAG
F3 UNS3 fw	GCACTGAAGGTCCTCAATCGCACTGGAAACATCAAGGTCGAGGCCAGGAACCGTAAAA
Cargo UNS5 rev	CTCTAACGGACTTGAGTGAGGTTGTAAAGGGAGTTGGCTCGGGACCCTCTGAACAAATCCAGATG
UNS5 fw	GAGCCAACTCCCTTTACAACCTCACTCAAGTCCGTTAGAG
For Gibson Assembly Sequencing	
VWB Int + UNS2	GAAGAGAAGCAGGACGAGC
VWB int + UNS3	TATGGAAAAACGCCAGCAAC
For Molecular Cloning Sequencing	
API4	GCCACATCGAAGCCGTCG
API6	TACAAGGGCTTCCAGGTG
PS1	AGGGCGGCGGATTTGTCC
Cargo UNS5 rev	CTCTAACGGACTTGAGTGAGGTTGTAAAGGGAGTTGGCTCGGGACCCTCTGAACa\AAATCCAGATG
For pSAM2 Integration Confirmation	
In *S. albus* genome	
SAM2 seq fw	CGGCCAGCTCTGCATCCC
SAM2 seq rev	CGGATTGTTTGCCGCCTTC
In Plasmid	
oriT seq fw	CTTTTCCGCTGCATAACCC
SAM2int seq rev	CACGTCGATGCACGCCTAC

Table 2. Sequences of primers used in this work.

4.2 Bacterial Transformation

E. coli Top 10

E. coli Top 10 was used for DNA cloning and DNA propagation.

E. coli Top 10 Preparation

Chemo competent *E. coli* Top10 (Invitrogen, USA) were prepared following the rubidium chloride method (38).

E. coli Top 10 Transformation

To transform Top 10 *E. coli*, an aliquot of previously prepared cells was thawed on ice. Once thawed, up to 20 µL of plasmid DNA was added to the tube and gently mixed. The mixture was then incubated on ice for 30 min followed by a 45 s heat shock at 42 ^{o}C. The mixture was then returned to ice for 1 min, after which time 1 mL of TSB media was added and the mixture was incubated at 37 ^{o}C for 1 h. Upon completion of the incubation period, the cells were suspended and spread onto 4 TSA plates containing hygromycin and X-gal. (35)

***S. albus* Transformation**

Final constructs were introduced into *S. albus* strains by protoplasts transformation.

Protoplast Preparation

10^8 colony forming units (CFU) of *S. albus* were first grown for 24 h in 10 mL of TSB media at 30 ºC and 250 rpm. Next, 750 µL of this culture was transferred to 25 mL of YEME 17% media and incubated for 36 h at 30 ºC and 250 rpm. After this period of incubation, the *S. albus* cells were pelleted and washed 3 times with 10.3% sucrose. After the final wash, the sucrose was removed and 10 mL of Medium P with 1 mg/ml of Lysozyme was added to digest the cell walls. This mixture was incubated at 30 ºC for 1 h, and then the cells were viewed under a microscope to confirm that the cell walls had been removed. The cells were then washed twice with Medium P to remove any residual lysozyme, filtered through sterile cotton filters to remove cell wall debris, and then stored in aliquots of 200 µL at -80 ºC (39).

Protoplast Transformation

To transform the protoplasts, 30 µL of plasmid DNA were added to an aliquot of protoplasts followed by the addition of 500 µL of polyethylene glycol 6000 1:3 in Medium P, to permeabilize the cell membranes. The cells were then spread on plates of supplemented R5 10% MgCl media that had dried for 2 h previously. The plates were then incubated at 30 ºC overnight. The following day hygromycin diluted into 1.5 mL of sterile water was spread over the plates for a final hygromycin concentration of 100 µg/mL in each plate. The plates were further incubated at 30 ºC until colonies were visible or up to 1 week. Visible colonies were then spread on Bennett plates containing 100 µg/mL of hygromycin to induce sporulation and remove potential false positive colonies (39).

4.3 Culture conditions

***E. coli* Top 10 Solid Culture**

Tryptic Soy Broth with agar (TSA) (35)

30 g/L of Tryptic Soy Broth mixture (Merck, Germany) with 22 g/L of agar in deionized water.

***E. coli* Top 10 Liquid Culture**

Tryptic Soy Broth (TSB) (35)

30 g/L of Tryptic Soy Broth mixture (Merck, Germany) in deionized water.

***S. albus* Growth for Protoplast Formation**

YEME 17% (39)

10 g/L of glucose, 3 g/L of yeast extract, 3 g/L of malt extract, 5 g/L of meat peptone, 170 g/L of sucrose, in deionized water. To 100 mL aliquots 3.75 mL of glycine 20%, and 200 µL of $MgCl_2$ 2.5 M were added.

***S. albus* Hyphal Growth**

Supplemented R5 Media 10% $MgCl_2$ (39)

103 g/L of sucrose, 0.25 g/L of K_2SO_4, 1.012 g/L of $MgCl_2$, 10 g/L of glucose, 0.1 g/L of hydrolyzed casein, 5 g/L of yeast extract, 5.73 g/L of TES, 2 mL/L of trace elements ($ZnCl_2$, $FeCl_3$, $CuCl_2$, $MnCl_2$, $Na_2B_4O_7$, $(NH_4)_6Mo_7O_{24}*4H_2O$), 22 g/L of agar, in deionized water. To 100 mL aliquots, 1 mL of KH_2PO_4 0.5%, 0.4 mL of $CaCl_2*2H_2O$ 5 M, 1.5 mL of L-Proline 20%, and 0.7 mL of NaOH 1 M were added.

***S. albus* sporulation**

Bennett Media (39)

10 g/L of glucose, 1 g/L of peptone, 1 g/L of yeast extract, 2 g/L of tryptone, and 20 g/L of agar in deionized water.

***S. albus* liquid culture for Naringenin Production**

NL333 Media (40)

5 g/L of glucose, 10 g/L of soluble starch, 10 g/L of malt extract, 3 g/L of yeast extract, 3 g/L of bactopeptone, 3 g/L of ammonium nitrate, 1 g/L of $CaCO_3$, in deionized water, pH 7.2

4.4 Naringenin Production, Extraction, Quantification

Spore Quantification

Once extensive sporulation was visible in the Bennett plate, the spores were scraped off the plate using a scalpel. They were then washed with sterile ultra-pure water and filtered through sterile cotton filters to remove hyphal growth. A serial dilution was then done from the spore mixture and the dilutions were plated on TSA plates with no

antibiotic. After 2 days, colonies were counted in whichever plate contained the most optimum density for counting colonies and in this way the colony forming units (CFU) / mL was calculated.

Liquid Culture

Liquid culture was done in glass 300 mL flasks containing invaginations around the base to prevent the bacteria from aggregating and clumping. 30 mL of NL333 media was used and spores were seeded at a density of 10^6 CFU/mL. The cultures were incubated at 30 ^oC and 200 rpm. Samples were collected from the cultures at 48 h and in the case of the multiple integration site experiment, 96 h as well.

Dry Biomass

To measure bacterial biomass in culture 1 mL of culture was drawn and centrifuged at maximum speed for two minutes to pellet the bacteria. The supernatant was removed, cells were washed twice in ultra-pure water, and then the clean pellet was left to dry at 60 ^oC for multiple days until completely dry. The pellets were then weighed in a precision balance.

Extraction

Flavonoids were extracted using an optimized organic extraction protocol. First, samples from culture were centrifuged and the supernatant was transferred to a separate tube from the pellet. The pellet underwent extraction first with one hour of incubation under agitation in acetone and then again in ethyl acetate. Separately, the supernatant underwent two rounds of liquid-liquid extraction with ethyl acetate. The pellet and supernatant extract fractions were pooled and dried down by vacuum centrifugation.

For the multiple integration site experiment, only the pellet underwent extraction using the same protocol as described above.

Quantification

Flavonoid quantification was done using optimized high performance liquid chromatography methods with diode array ultraviolet detection (HPLC-DAD). Dried extracts were resuspended in DMSO:Methanol 1:1 and 5 µL were injected into the HPLC. The samples were run on a gradient consisting of 10% buffer B for 5.44 min, increase to 35% buffer B over 16.33 min, 35% buffer B for 5.44 min, increase to 100% buffer B over 16.33 min, 100% buffer B for 11.46 min, decrease to 10% buffer B over 1 min, and 10%

buffer B for 5 min. Buffer A is HPLC grade water with 0.1% formic acid and buffer B is HPLC grade acetonitrile. The flow rate was 1 mL/min. The analytical column that was used was a Pursuit XR 5 C18 (Agilent, USA) with dimensions of 250x4.0 mm. The column temperature was 25 ^{o}C. The resulting chromatograms and UV spectra were analyzed using the HPLC Data Analysis software (Bruker, Germany). Naringenin standards (Sigma Aldrich, Spain) were previously run using the same method and using their chromatograms and known concentrations the naringenin peak could be identified in our chromatograms and the concentration in our samples could be calculated.

4.5 Integrase Integration Sites in *Streptomyces albus*

Three integrases were used in this work φC31, φBT1, and pSAM2. Details about the integration sites of these integrases can be viewed in Table 4.

Integrases		
Integrase	Integration Site	Gene Functions
ϕC31	XNR_3086 Gene	Chromosome condensation protein
ϕBT1	ZP_06586584 Gene	Integral membrane protein
pSAM2	Intergenic region between the XNR_2739 Gene and the XNR_2740 Gene	XNR_2739: Serine tRNA XNR_2740: Hypothetical Protein

Table 4. Integrases used in this work (41) (42) (43)

4.6 *Streptomyces albus* Strains

***S. albus* J1074**

S. albus J1074 is a strain of *S. albus* that has been modified to have defective *Sal*I restriction and modification enzymes (44). It is generally used as the wild-type strain in many research labs.

***S. albus* J1074 UOFLAV004**

The UOFLAV004 *S. albus* strain was generated previously in our lab. It has had BGCs coding for antimycin, candicidin, *epi*-alteramides, and paulomycins eliminated by CRISPR-Cas9 knockout. These BGCs compete with our naringenin production pathways for the usage of malonyl-CoA (45). The elimination of these BGCs increases the

availability of malonyl-CoA and thus this strain achieves greater flavonoid yields than the wild-type strain (data not yet published).

A full list of the strains that were developed and used in this project can be viewed in table 5.

Strains		
Name	Function	Origin
S. albus J1074	Served as a wild-type strain that had no integrations or further modifications.	Colección Española de Cultivos Tipo (CECT)
S. albus WT C15D pSAM2	*S. albus* J1074 with the pSAM2int hyg C15D plasmid integrated into the pSA2 integration site.	This project
S. albus WT NAR ϕC31	*S. albus* J1074 with the naringenin BGC integrated into the ϕC31 Integration site.	BIONUC, not yet published
S. albus WT NAR ϕBT1	*S. albus* J1074 with the naringenin BGC integrated into the ϕBT1 Integration site.	BIONUC, not yet published
S. albus WT NAR pSAM2	*S. albus* J1074 with the pSAM2int hyg cNARd plasmid integrated into the pSAM2 Integration site.	This project
S. albus UOFLAV004 NAR	*S. albus* J1074 with the UOFLAV04 modifications and the naringenin BGC integrated into the ϕC31 Integration site.	BIONUC, not yet published
S. albus UOFLAV004 NAR NAR	*S. albus* J1074 with the UOFLAV04 modifications and the naringenin BGC integrated into the ϕC31 and ϕBT1 Integration sites.	BIONUC, not yet published
S. albus UOFLAV004 NAR NAR NAR	*S. albus* J1074 with the UOFLAV04 modifications, the naringenin BGC integrated into the ϕC31 and ϕBT1 integration sites, and the pSAM2int hyg cNARd plasmid integrated into the pSAM2 Integration sites.	This project

Table 5. *S. albus* strains used in this work.

4.7 Data Analysis and Statistics

Figure generation and statistical analysis were done using the Prism software (GraphPad).

Statistical Analysis

For comparing naringenin production between different integration sites, a Kruskal-Wallis one-way ANOVA test with Dunn's multiple comparisons correction was used to generate significance values. For selecting the *S. albus* UOFLAV004 clone with the naringenin BGC integrated in three sites a Kolmogorov-Smirnov T test was used. For comparing *S. albus* growth rates between strains with different numbers of integrations and naringenin production between strains with different numbers of integrations, a two-way ANOVA test with a Tukey multiple comparison correction was used. To compare naringenin yields per mg of dry biomass at 96 hours a Kruskal-Wallis one-way ANOVA test with Dunn's multiple comparisons correction was used.

5. Results

5.1 Plasmid Assembly

5.1.1 pSAM2int hyg C15D Plasmid

The first task in this *TFM* was to assemble a plasmid containing the genes coding for the pSAM2 integrase and its *attP* site, the hygromycin resistance cassette, a molecular cloning gadget for a 4 gene BGC containing the *lacZ* operon, an origin of replication in *E. coli*, and genes to facilitate conjugation between *E. coli* and *S. albus*, all in the pSEVA architecture. SEVA stands for Standard European Vector Architecture (46). In this format, plasmids are divided into different regions, with each region joined by a common DNA sequence which are called UNS zones in our group. The plasmid regions consist of the antibiotic resistance cassette, the *Streptomyces* replication cassette, the *E. coli* replication cassette, and the gadget cassette. Because the regions are joined by common sequences, different regions from different plasmids can easily be joined to form new plasmids.

The pSAM2 integrase and hygromycin resistance genes were optimized and synthesized as described in Materials and Methods. The remaining regions were already available in other plasmids already in our laboratory's possession. Using PCR, the desired regions were amplified from each plasmid (Modular Cloning gadget and origins

of replication from Moclo lvl2 VWBint hyg C15D using Crisper F1 UNS2 rev and Crisper F1 UNS1 fw bis primers as well as Crisper F3 UNS3 fw and Cargo UNS5 rev primers, hygromycin resistance from pSEVA181 hygR using UNS5 fw and UNS1 rev primers, and pSAM2 integrase from pSEVA181 pSAM2 using UNS3 rev and UNS2 fw primers), purified by gel electrophoresis, and then joined by Gibson Assembly Figure 5. The resulting DNA product was transformed into competent *E. coli* cells and cultured in TSA media containing hygromycin and X-Gal. (Galactose bound to indole: bacteria containing the lacZ operon hydrolyze X-gal, which causes them to grow displaying a blue color in its the presence.)The resulting blue colonies were selected for further liquid culture as they had to be both resistant to hygromycin and contain the *lacZ* operon in the gadget to grow effectively and display their blue color. After further liquid culture, plasmid DNA was extracted and digested with *Sal*I restriction enzyme to confirm the correct assembly of the plasmid Figure 6. The pattern of clone 1 did not match the expected band pattern as simulated *in silico* while clones 2-4 did match the expected band pattern. Clone 2 was selected to continue and was sent for Sanger sequencing using the primers stated in Table 2, which confirmed its correct assembly. The complete plasmid can be seen in supplementary figure 1.

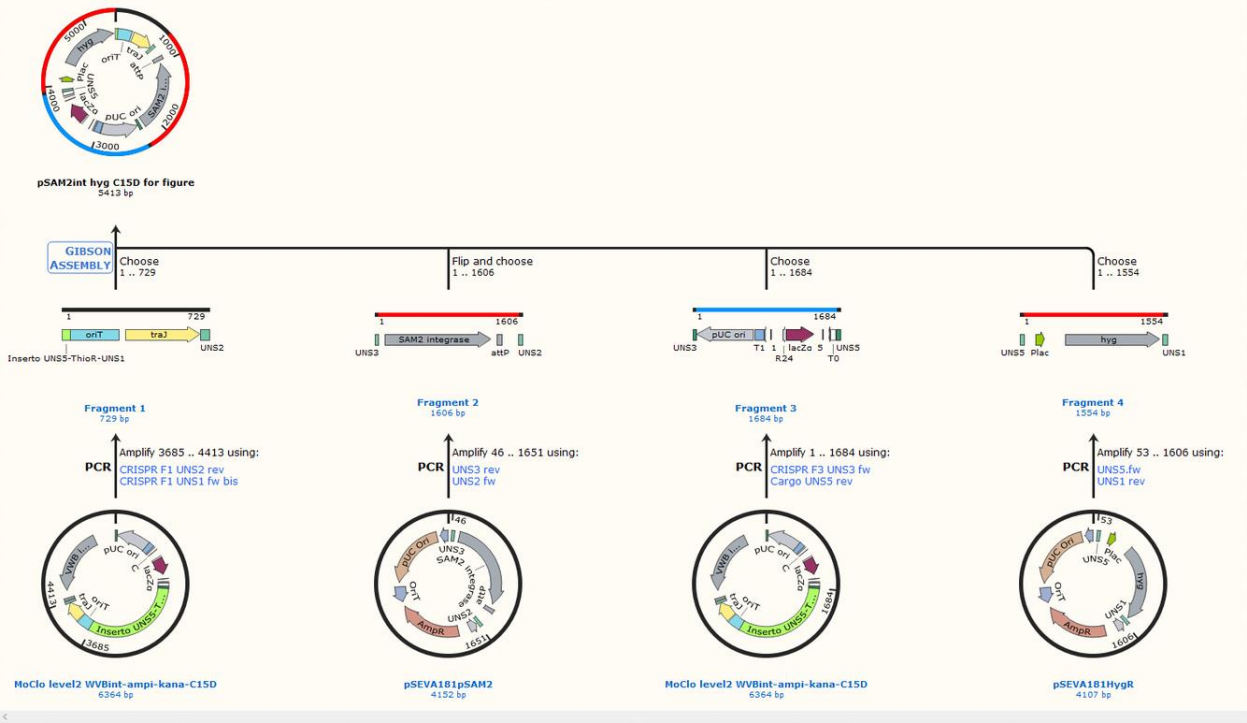


Figure 5. The four plasmids on the bottom represent the plasmids of origin of the regions in the new pSAM2int hyg C15D integrating plasmid. A description of each plasmid and where it came from can be found in table 1. The middle section displays the amplicons that were produced from each origin plasmid. The top section represents the final pSAM2 integrating plasmid assembled from the four amplicons. A detailed description of each plasmid can be found in Table 1.

A. **B.**

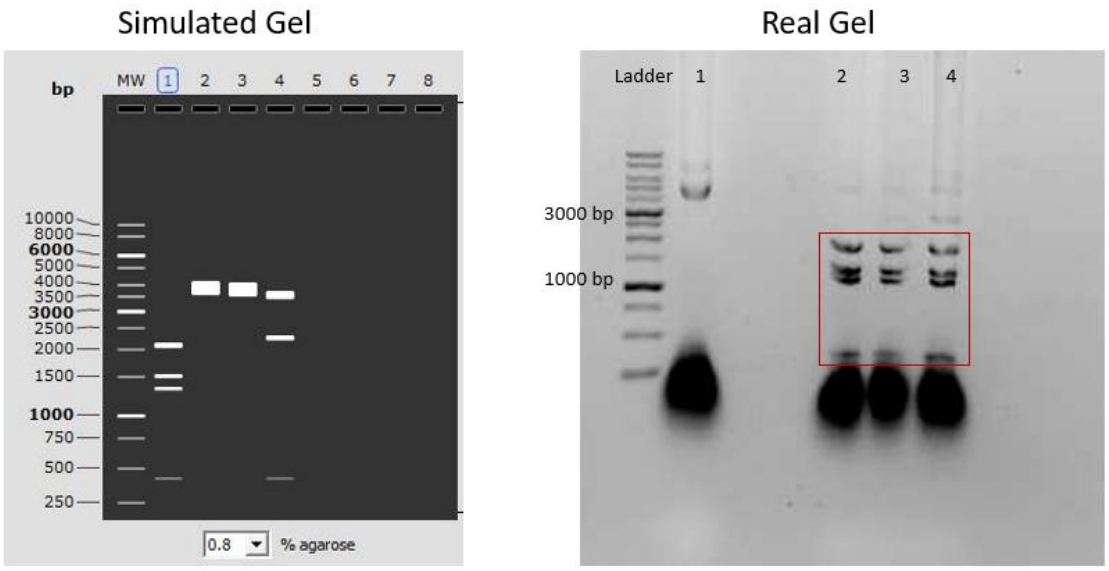


Figure 6. A: The simulated gel is an *in silico* simulation of what the plasmid should look like when digested with *Sal*I and run on a 0.7% agarose gel. Column 1 represents the correctly assembled pSAM2int hyg C15D plasmid, column 2 represents the pSEVA181 pSAM2 plasmid, column 3 represents the pSEVA181 hygR plasmid, and column 4 represents the MoClo lvl2 VWBint ampi kana C15D plasmid. B: The real gel represents the *Sal*I digestion of the plasmids that were extracted from the blue *E. coli* transformant colonies. The pattern of clone 1 does not match the *in silico* pattern, while the patterns of clones 2-4 do.

5.1.2 pSAM2int hyg NAR Plasmid

The second task in this TFM was to insert the naringenin BGC into the gadget in the newly assembled pSAM2int hyg C15D plasmid. The CIDAR modular cloning method was used to insert the naringenin BGC genes coding for tyrosine ammonia lyase, 4-coumaric acid-CoA ligase, chalcone synthase, and chalcone isomerase into the gadget Figure 7. The basis of the CIDAR modular cloning method is that plasmids containing modular parts can be joined in highly controlled orders to create higher order plasmids.

Level 0 plasmids contain single parts such as genes coding for enzymes, promotors, etc. Level 1 plasmids are made up with level 0 plasmids and contain all the necessary materials for the desired enzyme to be transcribed efficiently. Level 2 plasmids are made of level 1 plasmids and contain all the enzymes for a respective BGC. *Bbs*I and *Bsa*I restrictions enzymes are used to cut the plasmids on either side of their gadgets so that they can be joined in the new plasmid. Because these enzymes cut in locations that are adjacent to the sequences they detect, once a new gene is entered into the plasmid the enzyme recognition site is removed and the plasmid can no longer be cut. Because of this, the CIDAR MoClo method is highly efficient and specific.

pSEVABT1intApra1A12-permE-RiboJRES-TAL-ttsbiB is the level 1 plasmid from which the tyrosine ammonia lyase gene was excised. pSEVABT1intApra2AE3-sf14-RiboJRBS-4CL-ttsbiB is the level 1 plasmid from which the 4-coumaric acid-CoA ligase gene was excised. pSEVABT1intApra3AI4-Sp25-RiboJRBS-CHS-ttsbiB is the level 1 plasmid from which the chalcone synthase gene was excised. pSEVABT1intApra3AI4-Sp43-RiboJRBS-CHI-ttsbiB is the plasmid from which the chalcone isomerase gene was excised. These four plasmids were mixed along with the pSAM2int hyg C15D plasmid and were incubated following the CIDAR MoClo protocol. The mixture containing the product of the CIDAR MoClo protocol as well as any unreacted plasmid was then transformed into competent *E. coli* cells, which were cultured in TSA media with hygromycin and X-Gal. White colonies that grew in these conditions had to have resistance to hygromycin and had to have the LacZ operon in their gadget removed, indicating that they likely contained the correctly assembled plasmid with the naringenin BGC. Three white colonies were grown further in liquid culture, and the plasmids were then extracted, digested with *Sal*I restriction enzyme, and then run in an agarose gel to confirm proper plasmid formation Figure 8. All three clones displayed the correct band pattern as simulated *in silico*, but sample 1 also contained a larger band at 5 kb. This very large band is likely due to incomplete digestion of all the copies of the plasmid, but even so this clone was avoided for further processing. The bottom expected band (above 250 bp) is not visible in the experimental gel, likely because it is covered by the RNA blob in the same region. In the theoretical gel, the bands bellow 1500 bp and above 2000 bp are clearly pairs of two bands each. In the experimental gel, only one band is visible at each

of these positions, but this is likely due to insufficient resolution. Clone 2 was selected for further processing and its correct assembly was confirmed by Sanger sequencing using the primers described in Table 2. The complete plasmid can be seen in supplementary figure 2.

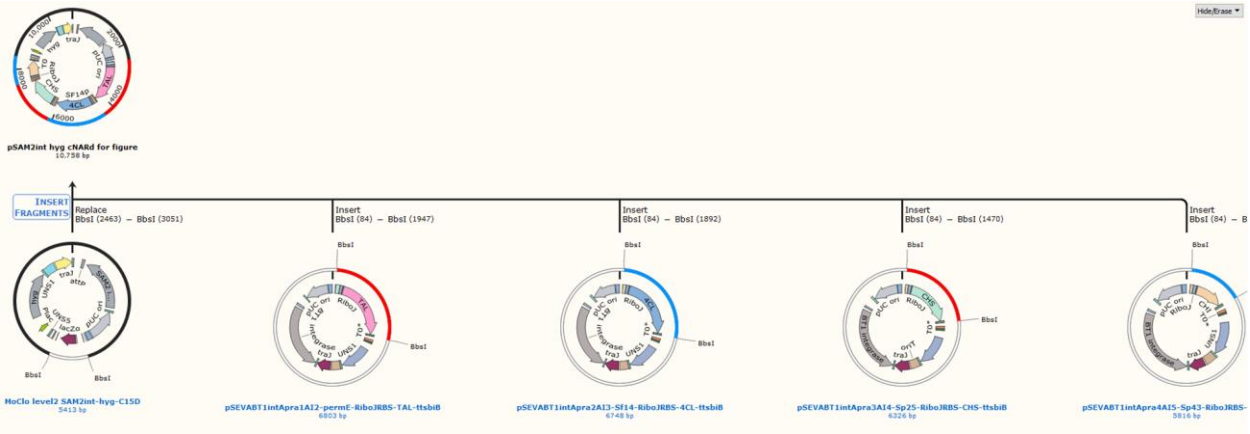
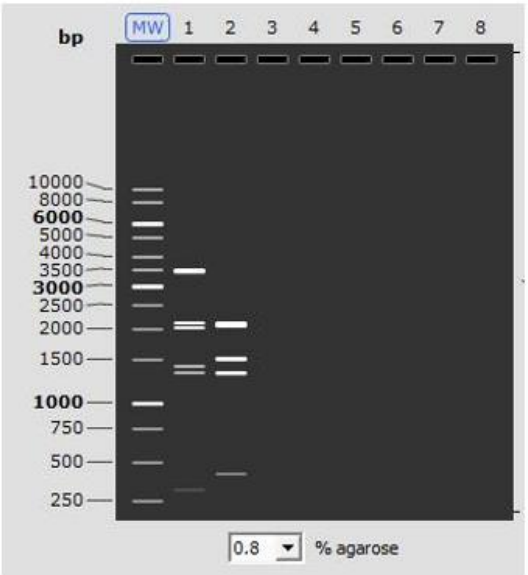


Figure 7. The 5 plasmids on the bottom represent the level 1 plasmids of origin of the 4 genes in the naringenin BGC as well as the level 2 pSAM2int hyg C15D plasmid on the left. The plasmid on the top represents the new pSAM2int hyg NAR plasmid with the naringenin BGC inserted into its molecular cloning gadget. A detailed description of each plasmid can be found in table 1.

A.



B.

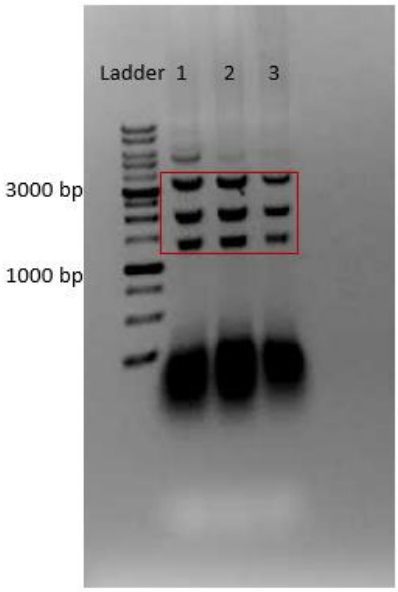


Figure 8. A: *In silico* theoretical *Sal*I digest gel. Column 1 represents the expected band pattern for the pSAM2int hyg NAR plasmid. Column 2 represents the expected band pattern for the pSAM2int hyg C15D plasmid. B: An electrophoresis gel of the *Sal*I digested plasmids that were extracted from the white colonies after transformation with the CIDAR MoClo reaction mixture. All three clones display the expected pSAM2int hyg NAR band pattern as predicted *in silico.*

5.2 Differences in Naringenin Production Between Different Integration Sites

As explained in the introduction, different regions in the bacterial genome experience different levels of transcriptional activity (33). Also, our group has previously demonstrated differences in naringenin production when using 2 previously established integrases, φC31 and φBT1 (data not yet published). Thus, a comparison of naringenin yields between strains with the naringenin BGC integrated into the φC31, φBT1, and the newly established pSAM2 integration sites was certainly warranted. To investigate these potential differences in naringenin production, we first confirmed the integration of the integrative plasmids that were generated in section 5.1. We then cultured strains containing the naringenin BGC integrated into each of the three integration sites that are currently in use by our group (φC31, φBT1, and pSAM2) and quantified their naringenin production.

5.2.1. *S. albus* WT NAR pSAM2 and *S. albus* WT C15D pSAM2 Strain Generation

Once the pSAM2 integrative plasmids were assembled, they were transformed into *S. albus* WT protoplasts as described in methods. This was done with both the pSAM2int hyg NAR plasmid as well as the pSAM2int hyg C15D plasmid to serve as a method control. Both plasmids successfully transformed *S. albus*, colonies formed in hygromycin containing media, and their spores were collected.

5.2.2. *S. albus* WT NAR pSAM2 and *S. albus* WT C15D pSAM2 Strain Verification

At this point, we knew that the plasmids could enter *S. albus* and confer resistance to hygromycin, but it was still necessary to confirm their integrative capabilities. Also, the

orientation by which the plasmid would integrate into the *S. albus* genome had not yet been demonstrated. For these reasons, we modified a method from Aubry et al. (32) which utilized PCR with primers that bound to the plasmid DNA on either side of the *attP* site and bound to the *S. albus* genomic DNA on either side of the *attB* site Figure 9. After transforming with our pSAM2 integrative plasmids, genomic DNA was extracted from the strains and multiple different combinations of the primers were used to create amplicons by PCR, which were then visualized through gel electrophoresis. The two combinations in which the primers from either DNA source were close enough to each other to produce an amplicon demonstrated the orientation by which the plasmid had entered the genome.

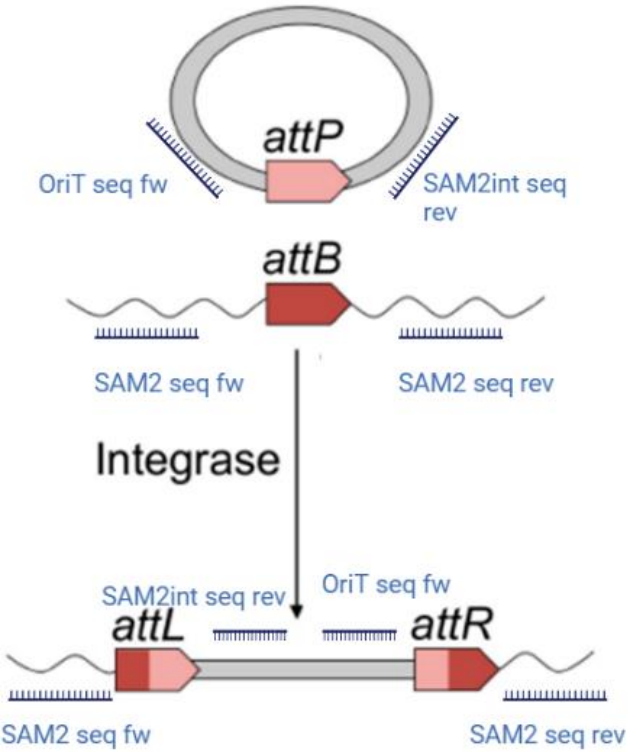


Figure 9. The plasmid with its *attP* site and the genome with its *attB* site, as well as the primers that bound to either side of the sites are shown. First, the plasmid and genome are separate, second, the plasmid is integrated into the genome. Adapted from Merrick and Zhao, 2018.

This technique was used using four combinations of primers (A, B, C, and D) on two *S. albus* WT C15D pSAM2 clones and on two *S. albus* WT NAR pSAM2 clones, as well as on wild-type *S. albus* to control for nonspecific amplification figure 10. Primer combo A used SAM2 seq fw and OriT seq fw. Primer combo B used SAM2 seq fw and SAM2int seq rev. Primer combo C used SAM2 seq rev and OriT seq fw. Primer combo D used

SAM2 seq rev and SAM2int seq rev (all primer sequences are available in Materials and Methods). Primer combos B and C, using the SAM2 seq fw-SAM2int seq rev and SAM2 seq rev-OriT seq fw primers created amplicons in both sets of clones while Primer pair A and D did not yield any clear bands in either of the samples. This indicates that the integrase inserts the plasmid with the B-C orientation, which amounts to the order pSAM2 integrase, pUC Ori, molecular cloning gadget, hygromycin resistance, OriT Supplemental Figure 3. The correct integration site was confirmed to be the same as what has been previously published (intergenic region between XNR_2739 and XNR_2740 genes, position 6,841,649 Supplementary Figure 4) by doing Sanger sequencing of the amplicon bands using the same primers and aligning with both the *S. albus* genome and the pSAM2 integrative plasmids.

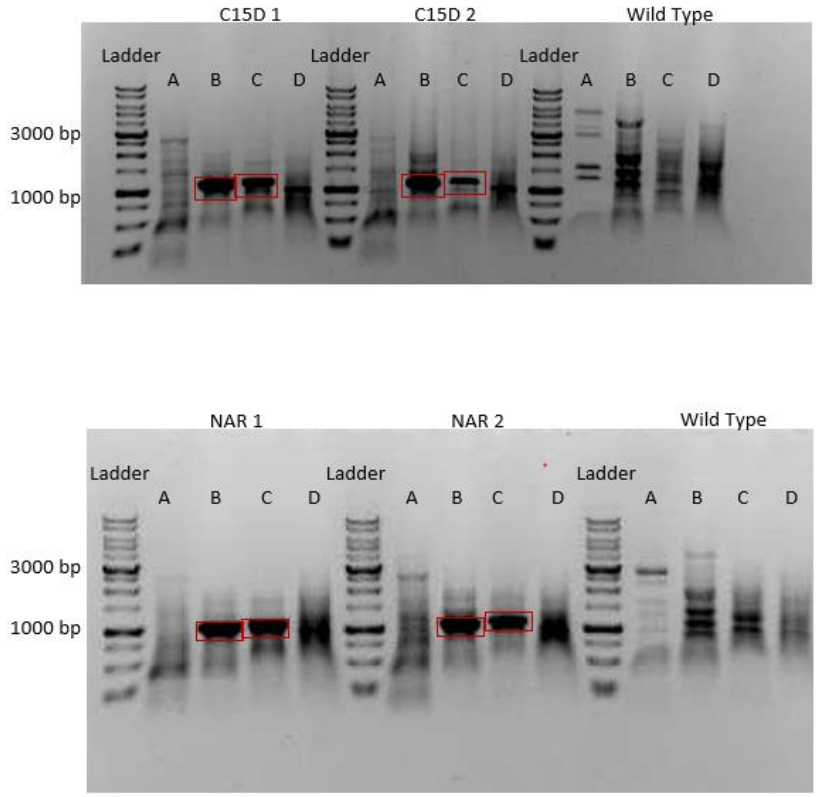


Figure 10. Electrophoresis gel of PCR amplicons using different primer combinations in 2 clones integrated with the pSAM2int hyg C15D plasmid (C15D 1 and 2), 2 clones integrated with the pSAM2int hyg NAR plasmid (NAR 1 and 2), and the wild-type *S. albus* strain. Column A is the product of a PCR with the primer pair SAM2 seq fw – OriT seq fw. Column B is the product of a PCR with the primer pair SAM2

seq fw – SAM2int seq rev. Column C is the product of a PCR with the primer pair SAM2 seq rev – OriT seq fw. Column D is the product of a PCR with the primer pair SAM2 seq rev- SAM2int seq rev.

Once the integration confirmation method was demonstrated to work and the correct primer pairs were selected, we next used the method to confirm integration of the pSAM2int hyg NAR plasmid into all the strains that were generated Figure 11. 4 Wild-type *S. albus* clones with the pSAM2int hyg NAR plasmid were generated. One wild-type *S. albus* strain with the pSAM2int hyg C15D plasmid and a wild-type *S. albus* strain with no integrated plasmids were used as controls. All four of the pSAM2int hyg NAR containing clones displayed the proper band pattern via gel electrophoresis. The clone containing the pSAM2int hyg C15D plasmid confirmed the expected band pattern and the lack of clear bands in the wild-type strain confirmed the lack of nonspecific amplification. The strains containing the naringenin BGC integrated into φC31 and φBT1 were not tested with the integration confirmation method since this had already been confirmed previously (data not yet published).

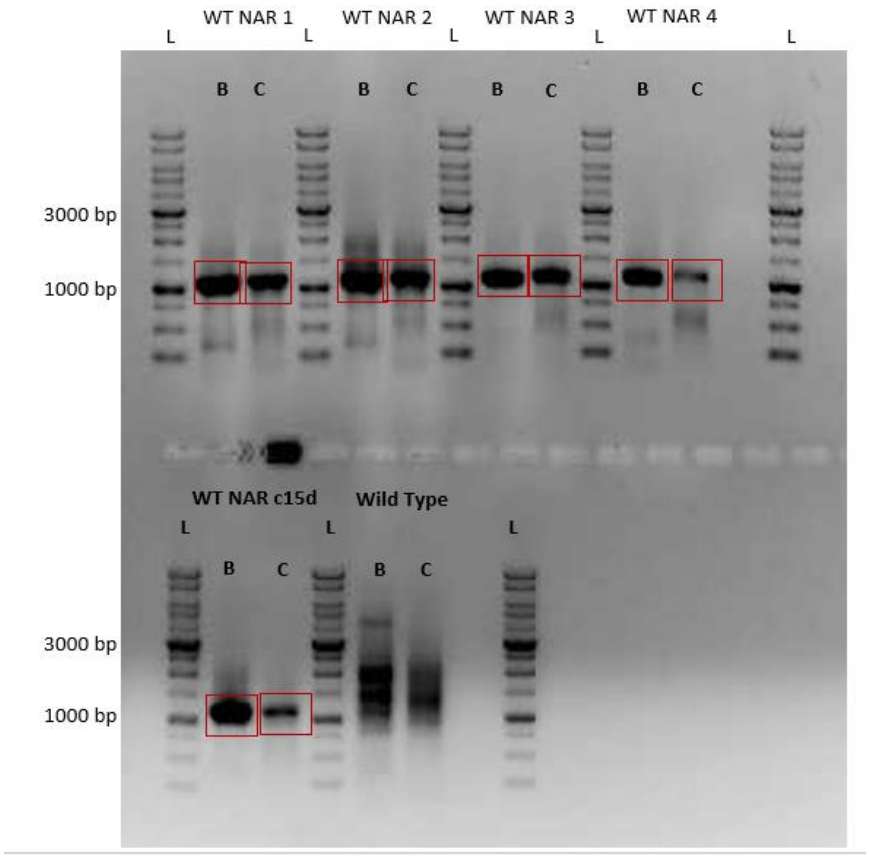
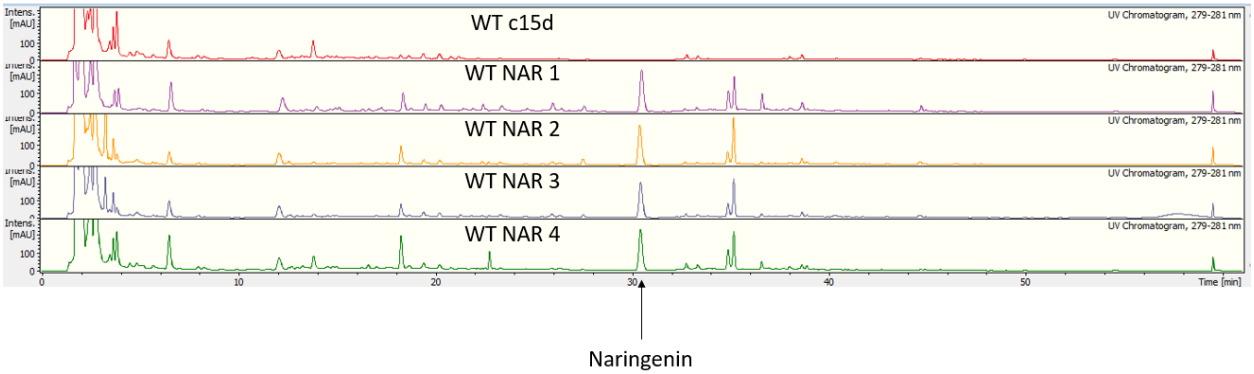


Figure 11. WT NAR 1-4 are the 4 *S. albus* WT NAR pSAM2 clones that were transformed with the pSAM2int hyg NAR integrating plasmid containing the naringenin BGC. WT NAR C15D is the *S. albus* clone that was transformed with the pSAM2int hyg C15D integrating plasmid that did not contain the naringenin BGC. Wild Type is the wild-type strain of *S. albus*, with no plasmids added. L is the standard molecular weight ladder. B and C are the primer pairs that confirm integration and orientation of integration of the plasmid.

5.2.3. Selection of a Representative Clone of *S. albus* WT NAR pSAM2

Once it was confirmed that the 4 clones contained the pSAM2int hyg cNARd plasmid integrated into the pSAM2 integration site, it was necessary to establish an average level of naringenin production to be able to select a representative strain. Although we confirmed the integration of our plasmid into the host genome, we did not sequence the entire genome and thus cannot rule out the presence of unexpected mutations. As such, variation in flavonoid production often exists between samples and without establishing an average yield for the strain it is impossible to know what differences in production are caused by your modification and which are caused by unexpected mutations. For this reason, we cultured all 4 *S. albus* WT NAR pSAM2 clones along with 1 *S. albus* WT C15D pSAM2 clone in liquid culture, extracted the flavonoids as described in methods after 48 h of culture, and quantified the naringenin in each sample by HPLC. The 5 chromatograms were compared and showed clear differences between the *S. albus* WT NAR pSAM2 samples and the *S. albus* WT C15D pSAM2 sample. Between the *S. albus* WT NAR pSAM2 samples, no major variation was observed Figure 12A. A clear peak was observed in the *S. albus* WT NAR pSAM2 samples and not in the *S. albus* WT C15D pSAM2 sample at 31 min. This retention time is known to be the naringenin retention time in our HPLC method (described above). Also, the UV spectrum at this peak showed maximum absorbance at 290 nm, which is known to be the naringenin UV absorbance spectrum Figure 12B. *S. albus* WT NAR pSAM2 sample 4 produces a peak at 24 min that is not produced in the other *S. albus* WT NAR pSAM2 samples. There is a lot of variation in the size of the peak that elutes at 18 min, with sample 4 producing by far the largest peak. *S. albus* WT NAR pSAM2 sample 1 produces peaks in the 20-30 min range that are not produced by the other *S. albus* WT NAR pSAM2 samples.

A.



B.

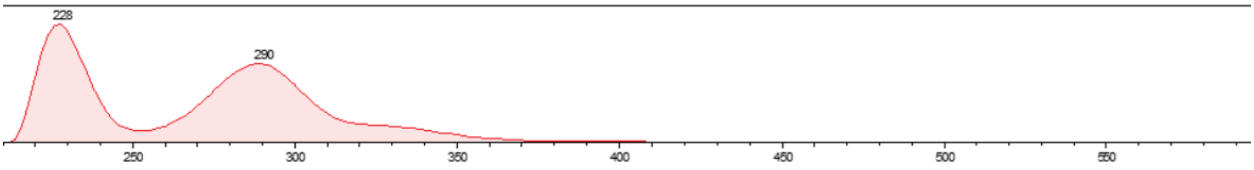


Figure 12. A: WT C15D is the *S. albus* WT C15D pSAM2 strain. WT NAR1-4 are the four *S. albus* WT NAR pSAM2 clones. All four naringenin BGC samples show a clear peak at 31 min, which aligns with the naringenin standard. B: UV absorbance spectrum of the 31-min peak. 290 nm aligns with the known naringenin maximum UV absorbance spectrum. 228 nm is the known absorbance of acetonitrile, which is present in our HPLC buffer B.

In addition to comparing the chromatograms, the area of the naringenin peaks was also compared to those generated from a naringenin standard HPLC run for quantification Figure 13. Because the samples were not cultivated in replicate, the statistical significance of the differences between the samples could not be calculated. Even so, the average production between the 4 *S. albus* WT NAR pSAM2 samples was 2.44 mg/L, with sample 4 clearly above the average and sample 2 clearly below the average. Samples 1 and 3 were the most representative from the naringenin production perspective. Since sample 1 produced peaks in the chromatogram that the other four did not produce, sample 3 was declared the most representative.

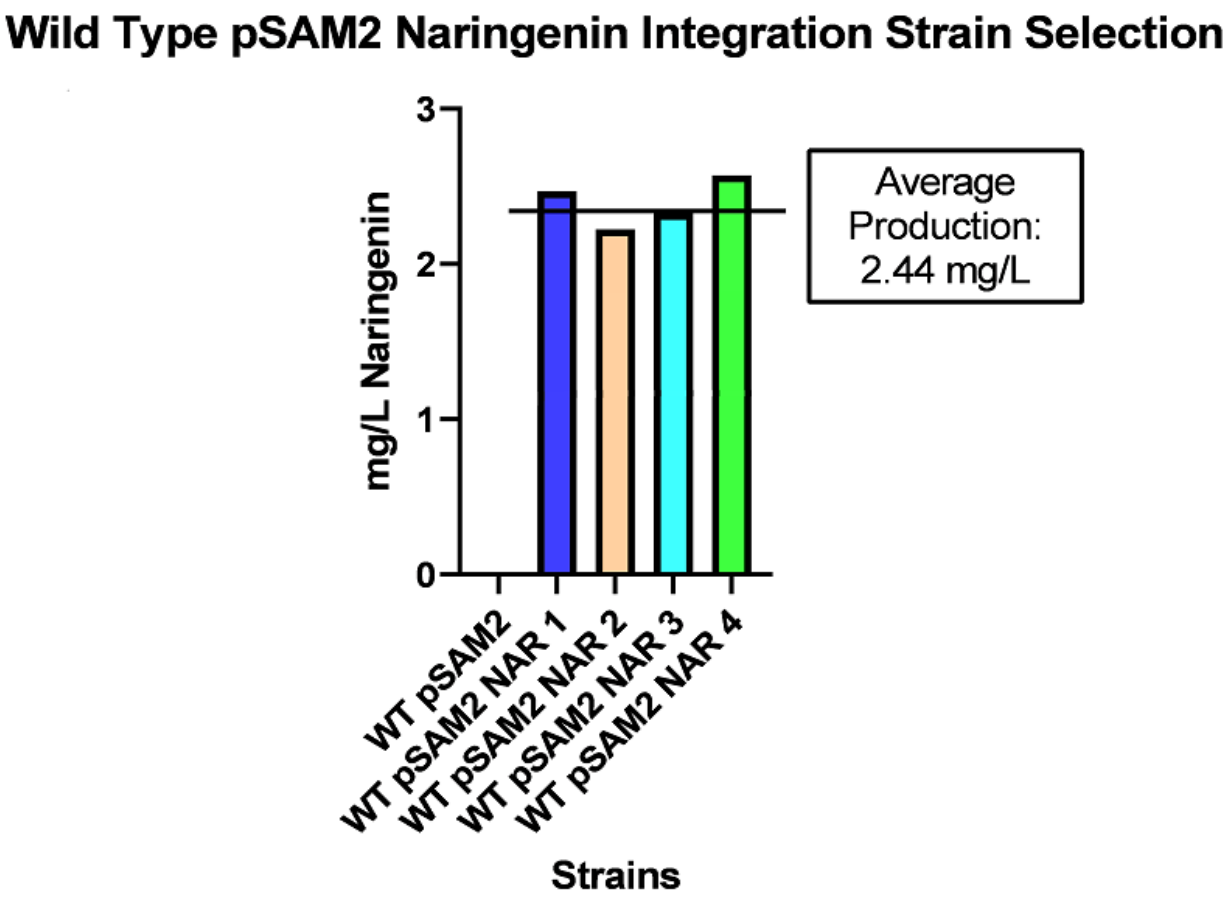


Figure 13. Naringenin production by different clones of the same *S. albus* WT NAR pSAM2 strain. WT pSAM2 is the *S. albus* WT C15D pSAM2 strain that was transformed with the pSAM2int hyg C15D plasmid. WT pSAM2 NAR1-4 are the four clones of *S. albus* WT NAR pSAM2 that were transformed with the pSAM2int hyg NAR plasmid. Samples pSAM2int NAR 1 and 3 are closest to the mean naringenin yield from this group of samples.

5.2.4. Cultures for Comparison of Naringenin Production

Once a representative clone was selected (clone 3), we compared naringenin yields between wild-type strains containing the naringenin BGC integrated into three different locations of the *S. albus* genome (*S. albus* WT NAR φC31, *S. albus* WT NAR φBT1, and *S. albus* WT NAR pSAM2). Naringenin yields were measured after 48 h of culture (using 3 replicates per experiment) Figure 14. Of the three strains, *S. albus* WT NAR pSAM2 produced the largest naringenin yield with an average of 2.769 mg/L, followed *S.*

albus WT NAR φBT1 with an average of 2.350 mg/L, and *S. albus* WT NAR φC31 produced the least with an average yield of 2.192 mg/L. The difference in yield between *S. albus* WT NAR pSAM2 and *S. albus* WT NAR φC31 was statistically significant as per a Kruskal-Wallis one-way ANOVA test and Dunn's multiple comparisons correction, with an adjusted p value of 0.0219. The differences between φC31 and φBT1 and between pSAM2 and φBT1 were not statistically significant.

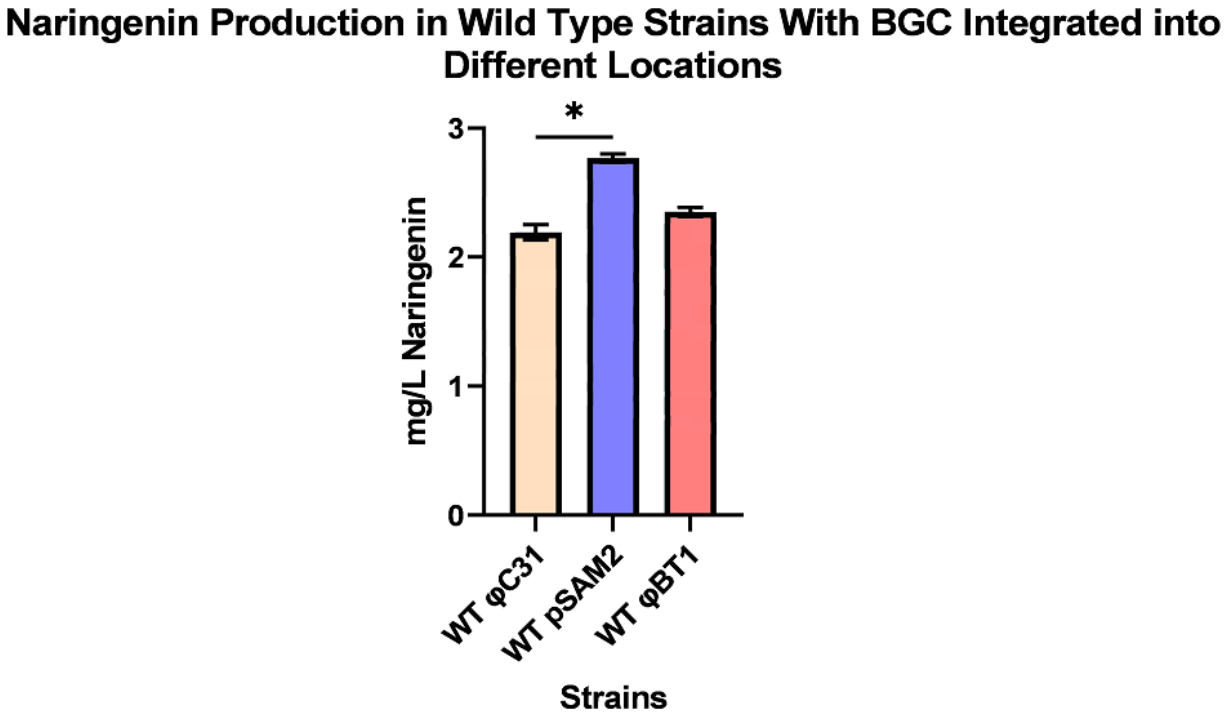


Figure 14. Naringenin production by *S. albus* WT strains with naringenin BGC integrated into different chromosomal sites. φC31 is the *S. albus* WT NAR φC31 strain with the naringenin BGC integrated into the φC31 integration site. WT pSAM2 is the *S. albus* WT NAR pSAM2 strain with the naringenin BGC integrated into the pSAM2 integration site. φBT1 is the S. albus WT NAR φBT1 strain with the naringenin BGC integrated into the φBT1 integration site. The asterisk and line indicate statistical significance between those samples.

5.3 Improving Naringenin Yield by Integrating the Naringenin BGC into Multiple Sites in the Same *S. albus* Strain

A second question that we wished to investigate was if the integration of the same BGC in multiple sites in the same strain could increase natural product yield. Our lab has previously demonstrated that integrating the naringenin BGC into both the φC31 and φBT1 sites in the same *S. albus* UOFLAV004 strain significantly increased naringenin yield compared to a similar *S. albus* UOFLAV004 strain with the naringenin BGC integrated into only the φC31 site (data not yet published). It is hypothesized that increasing the number of copies of the naringenin BGC in the same strain increases the amount of the enzymes encoded by the BGC, which thus increases the amount of naringenin producing enzymatic reactions and increases yield. To see if integrating the naringenin BGC into a third position in the same strain could further increase yield, we took the *S. albus* UOFLAV004 strain that already had the naringenin BGC integrated into the φC31 and φBT1 sites (*S. albus* UOFLAV004 NAR NAR) and used the newly assembled pSAM2int hyg NAR integrative plasmid to integrate the naringenin BGC into the pSAM2 site.

5.3.1 *S. albus* UOFLAV004 NAR NAR NAR Strain Generation

The *S. albus* UOFLAV004 NAR NAR strain was transformed with the pSAM2int hyg NAR plasmid to generate the *S. albus* UOFLAV004 NAR NAR NAR strain, as described in Materials and Methods. After transforming, only 2 colonies grew in the Bennett media plate, demonstrating the difficulty of inserting so many plasmids into the bacterial genome.

5.3.2 S. albus UOFLAV004 NAR NAR NAR Strain Verification

Both generated clones underwent the integration confirmation PCR assay to confirm that the plasmid had integrated into the pSAM2 site Figure 15. The assay was conducted in the same way as in section 5.2.2. Clones 1 and 2 both display the desired 1 kb bands with both sets of primers, indicating that the pSAM2int hyg NAR plasmid did integrate in both samples. No clear bands are present in the *S. albus* UOFLAV004 NAR NAR strain, indicating that no nonspecific amplification occurred.

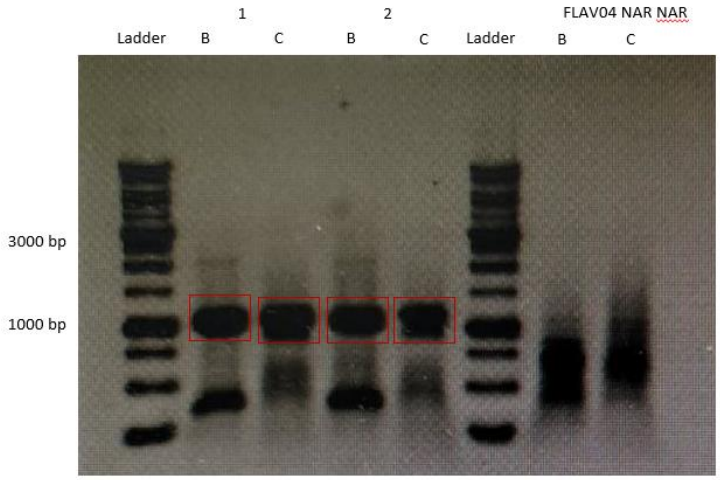


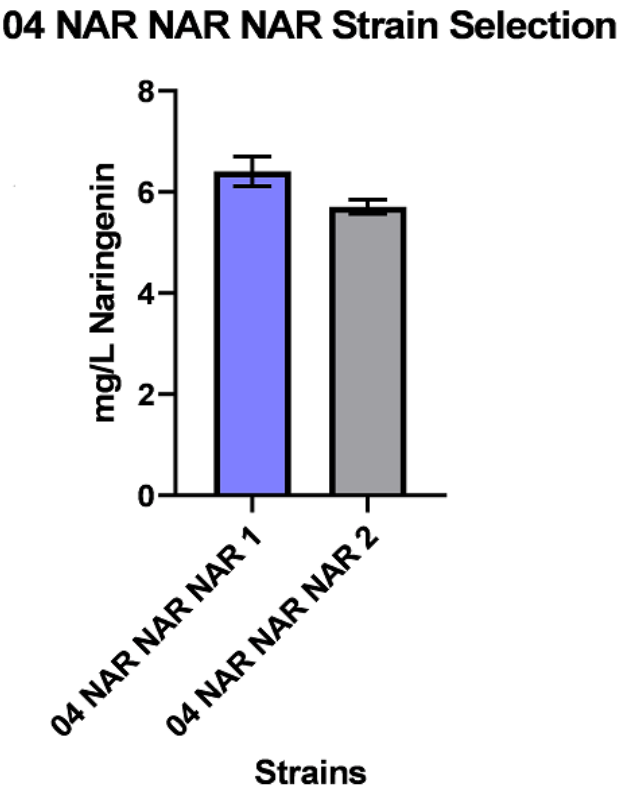
Figure 15. Agarose gel showing PCR amplifications of the pSAM2int integration site in 2 *S. albus* UOFLAV004 NAR NAR NAR clones. 1 and 2 are the two *S. albus* UOFLAV004 NAR NAR NAR clones that appeared after transformation with the pSAM2 integrative plasmid. FLAV004 NAR NAR is the *S. albus* UOFLAV004 NAR NAR strain with the naringenin BGC integrated into the φC31 and φBT1 sites. B and C are the primer pairs that were used to make the amplicon, as described section 5.2.2

5.3.3 *S. albus* UOFLAV004 NAR NAR NAR Representative Clone Selection

Once the *S. albus* UOFLAV004 NAR NAR NAR strain was generated and integration was confirmed, it was necessary to establish an average naringenin yield to confirm that the changes in naringenin yield were caused by the third integration and not by unexpected mutations. Ideally, multiple clones would have been compared as was done in section 5.2.3. Unfortunately, the *S. albus* UOFLAV004 strain does not grow as well as the wild-type strain, and with each successive integration, the strain becomes more difficult to work with. For these reasons, despite our best efforts, only two *S. albus*

UOFLAV004 NAR NAR NAR clones were generated. With only two clones, an average yield could not be established. For this reason, each clone was cultured in duplicate so that a Kolmogorov-Smirnov T test could be used to establish whether the difference in yield between the two clones was significant or not. If no significant difference in yield was found between the two clones, it would be safe to assume that both were representative samples since the probability of unexpected mutations causing the same alteration in yield in both samples is exceedingly low. If significant differences in yield were detected between both clones, then further clones would need to be made to establish an average yield. Clone 1 produced an average yield of 6.410 mg/L and clone 2 produced an average yield of 5.708 mg/L but the difference between the two clones had a p value of 0.3333 and thus no significant difference in yield was detected. Also, both clones showed remarkable similarity in their HPLC chromatograms Figure 16. For this reason, both clones were declared representative and clone 1 was selected to be used for further experimentation.

A.



B.

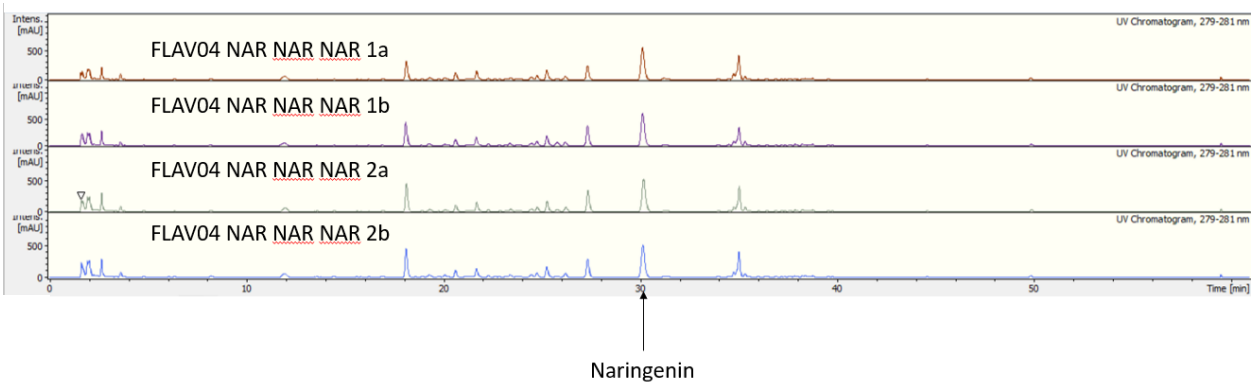


Figure 16. A: Naringenin yields from two *S. albus* UOFLAV004 NAR NAR NAR clones with the naringenin BGC integrated into three sites (φC31, φBT1, pSAM2). Both strains were cultured in duplicate and no significant difference in naringenin yield was detected between the clones. B: HPLC chromatograms of organic extracts. FLAV04 NAR NAR NAR is the *S. albus* UOFLAV004 NAR NAR NAR strain. 1a and 1b are replicate cultures from clone 1. 2a and 2b are replicate cultures from clone 2. All four chromatograms look remarkably similar, with no easily observable differences.

5.3.4 Comparison of Naringenin Production Between *S. albus* UOFLAV004 NAR, *S. albus* UOFLAV004 NAR NAR, and *S. albus* UOFLAV004 NAR NAR NAR

With a representative *S. albus* UOFLAV004 NAR NAR NAR clone selected, we next set out to compare naringenin production between it and its predecessor strains which only had the naringenin BGC integrated into one (φC31) or two (φC31 and φBT1) sites (*S. albus* UOFLAV004 NAR and *S. albus* UOFLAV004 NAR NAR). All three strains were cultured in triplicate and both biomass and naringenin yield were measured at 48 and 96 h.

Dry biomass was measured at both 48 and 96 h to investigate whether the differential integrations effected the *S. albus* UOFLAV004's growth rate Figure 17. *S. albus* UOFLAV004 NAR grew to an average of 6.33 mg/mL at 48 h and 4.76 mg/mL at 96 h. *S. albus* UOFLAV004 NAR NAR grew to an average of 5.57 mg/mL at 48 h and 5.43 mg/mL at 96 h. *S. albus* UOFLAV004 NAR NAR NAR grew to an average of 7.37 mg/mL at 48 h and 5.07 mg/mL at 96 h. For all three samples biomass was greater at 48 h than at 96 h, indicating a possible depletion of nutrients and subsequent die-off, which is also observed in previous data by our group (data not yet published). To compare the three samples at

each time point a two way ANOVA test was used with a Tukey's multiple comparison correction. No significant differences in growth at each time point were detected between the strains.

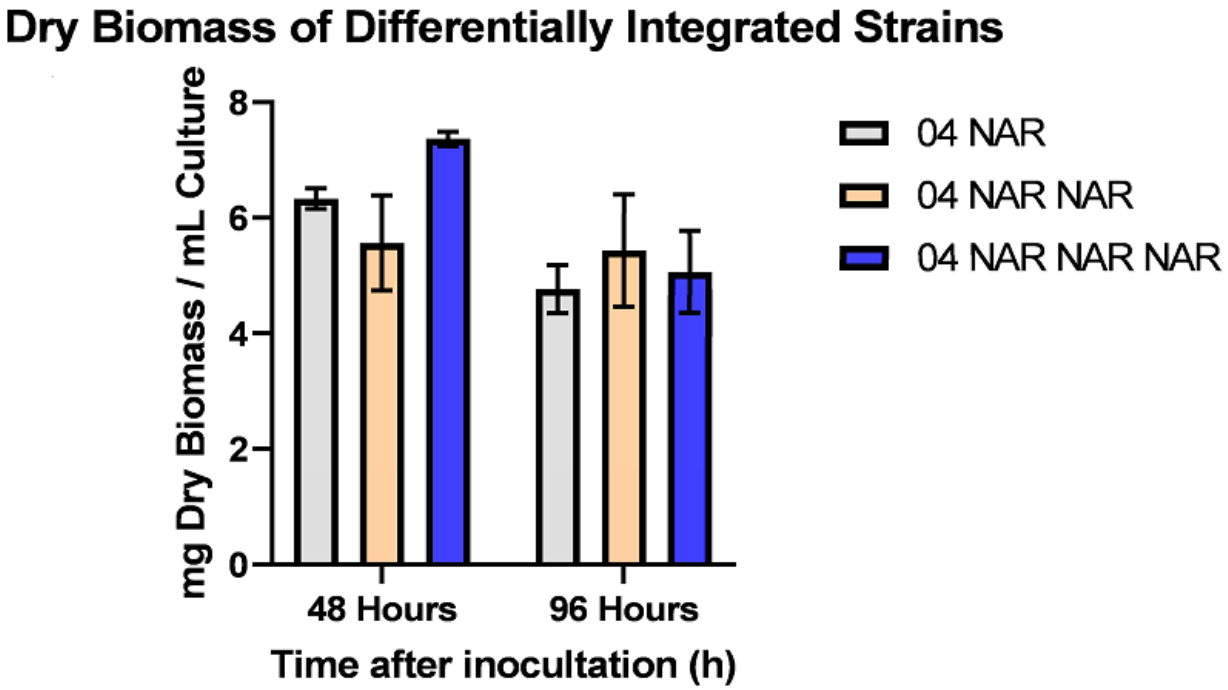


Figure 17. Dry biomass of *S. albus* UOFLAV004 NAR, *S. albus* UOFLAV004 NAR NAR, and *S. albus* UOFLAV004 NAR NAR NAR at 48 and 96 h. 04 NAR is the *S. albus* UOFLAV004 NAR strain. 04 NAR NAR is the *S. albus* UOFLAV004 NAR NAR strain. 04 NAR NAR NAR is the *S. albus* UOFLAV004 NAR NAR NAR strain.

Naringenin production was also measured at both 48 and 96 h to investigate whether the differential integrations increased naringenin yields Figure 18. *S. albus* UOFLAV004 NAR produced an average of 1.464 mg/L and 2.043 mg/L of naringenin at 48 h and 96 h respectively. *S. albus* UOFLAV004 NAR NAR produced an average of 2.359 mg/L and 3.789 mg/L of naringenin at 48 h and 96 h, respectively. *S. albus* UOFLAV004 NAR NAR NAR produced and average of 2.828 mg/L and 4.635 mg/L of naringenin at 48 h and 96 h, respectively. Naringenin production was clearly higher at 96 h than at 48 h, following an inverse pattern to that of the dry biomass. This has also been found in previous experiments by our group (data not yet published). To compare the three samples at each time point a two way ANOVA test was used with a Tukey's multiple comparison correction. At both time points *S. albus* UOFLAV004 NAR produced the

smallest naringenin yields, *S. albus* UOFLAV004 NAR NAR produced a middle amount, and *S. albus* UOFLAV004 NAR NAR NAR produced the most. At 48 h, the difference in naringenin production between *S. albus* UOFLAV004 NAR and *S. albus* UOFLAV004 NAR NAR NAR was statistically significant, with an adjusted *p*-value of 0.0333, while the yield from *S. albus* UOFLAV004 NAR NAR was not statistically significant from either of the other two. At 96 h the difference between *S. albus* UOFLAV004 NAR and *S. albus* UOFLAV004 NAR NAR, as well as the difference between *S. albus* UOFLAV004 NAR and *S. albus* UOFLAV004 NAR NAR NAR were both statistically significant, with adjusted *p*-values of 0.0078 and 0.0004 respectively.

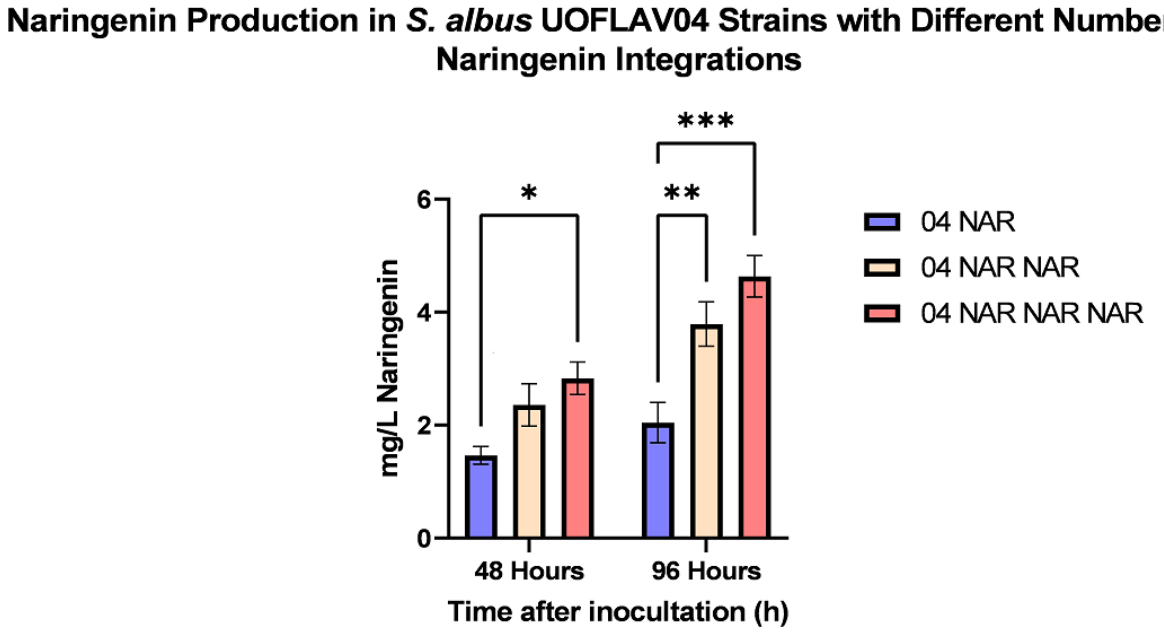


Figure 18. Naringenin yields from three strains with differing numbers of naringenin BGC integrations at 48 and 96 hours of culture. 04 NAR is *S. albus* UOFLAV004 NAR. 04 NAR NAR is *S. albus* UOFLAV004 NAR NAR. 04 NAR NAR NAR is *S. albus* UOFLAV004 NAR NAR NAR. Lines with asterisks represent statistical significance (P value < 0.05) between the two samples on either side of the line. A larger number of asterisks represents a smaller *p*-value and greater significance.

We then adjusted each strain's naringenin yields to its respective dry biomass to compare mg of naringenin production per mg of dry biomass. Here we found that, at 96 h of culture, *S. albus* UOFLAV004 NAR produced an average of 0.429 mg naringenin / mg dry biomass. *S. albus* UOFLAV004 NAR NAR produced 0.697 mg naringenin / mg dry biomass. *S. albus* UOFLAV004 NAR NAR NAR produced 0.915 mg naringenin / mg

dry biomass Figure 19. This pattern clearly followed the same pattern as the total naringenin yields (figure 18). To compare the samples, a Kruskal-Wallis two-way ANOVA test was used along with a Dunn's multiple comparison correction test. Due to significant variation between replicates, no statistical differences were detected between the three samples.

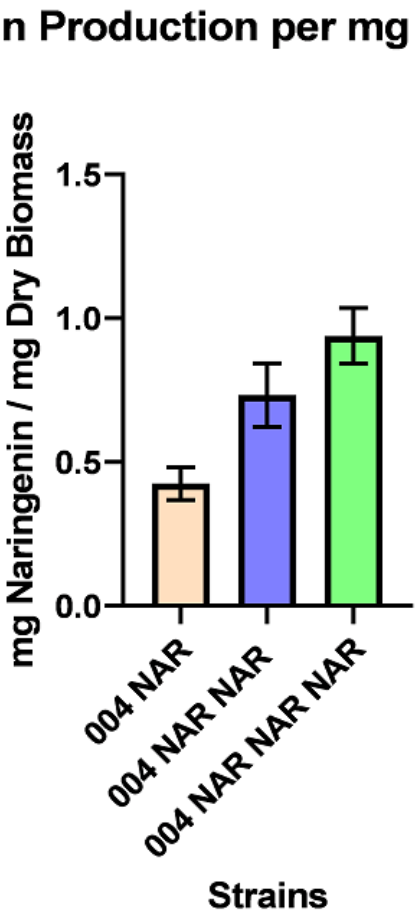


Figure 19. A comparison of naringenin yields per mg of dry biomass for each sample at 96 hours of culture. 004 NAR is *S. albus* UOFLAV004 NAR. 004 NAR NAR is *S. albus* UOFLAV004 NAR NAR. 004 NAR NAR NAR is *S. albus* UOFLAV004 NAR NAR NAR. No statistical significance was detected between the samples.

6. Discussion

6.2 Plasmid Assembly and Integration Functionality

In this work, we generated an integrative plasmid in the pSEVA architecture that contained the pSAM2 integrase, hygromycin resistance genes, and a modular cloning gadget for insertion of four genes, and then confirmed its ability to integrate into the *S. albus* genome (pSAM2int hyg C15D). We then inserted the four genes coding for the naringenin biosynthetic gene cluster into the gadget (pSAM2int hyg NAR) and used the plasmid to transform and stably integrate the genes into the genome of *S. albus* J1074 (WT) and *S. albus* UOFLAV004. The generation of this plasmid allows our group to insert genes into a third location of the *S. albus* genome (φC31, φBT1, and now pSAM2). With this functionality, our group will be able to generate more complex metabolic biosynthesis pathways, increasing the variety of potentially synthesizable flavonoids, as well as their production yields.

6.3 Comparison of Naringenin Production Between Different Integration Sites

We successfully demonstrated that naringenin production varies depending on which integration site its BGC is integrated into. The difference in naringenin yield between *S. albus* WT NAR φC31 and *S. albus* WT NAR pSAM2 is from 2.192 mg/L to 2.769 mg/L, a 1.26x increase. Also, *S. albus* WT φBT1 produces more naringenin than *S. albus* WT φC31 and less then *S. albus* WT pSAM2, although these differences are not statistically significant. We hypothesize that these differences in naringenin yield are likely due to differing levels of transcriptional activity in different areas of the *S. albus* genome. Further experiments are required to fully explain why this phenomenon occurs. In the future, quantitative PCRs of the genes in the naringenin BGC could be conducted to measure levels of transcriptional activity at each integration site. Furthermore, exploring

this phenomenon with production of other natural products would confirm whether it is naringenin specific.

Regardless as to why this phenomenon occurs, its presence can have potential in increasing natural product yield. When integrating multiple different BGCs into the same strain, ensuring that the BGCs coding for precursor production are integrated into the most active sites and the BGCs coding for final product production are integrated into less active sites could help to prevent bottlenecks and drive the overall reaction dynamics forward, potentially improving final yields. Knowing integration site activity could have great implications in rational metabolic pathway design, but further experiments are necessary.

With this said, the differences in production between the three integration sites that we tested were not particularly large. Further development of other integration sites is necessary to confirm whether this is true more broadly. While selecting integration sites based on their activity may be helpful for rationally designing complex metabolic pathways, simply inserting a BGC into a more active site is not likely to produce drastically more final product. Thus, for simple metabolic pathways, other methods to increase production should be prioritized.

Detailed information about the integration sites used in this work can be found in Table 4, section 4.5.

6.4 Comparison of Naringenin Production with Differential Levels of Integration of the Same Biosynthetic Gene Cluster

We first demonstrated that increasing the number of copies of naringenin BGC containing plasmids integrated into the *S. albus* genome did not reduce the viability or growth rate of the organism. This indicates that the differences in naringenin yield between the three strains tested here (*S. albus* UOFLAV004 NAR, *S. albus* UOFLAV004 NAR NAR, and *S. albus* UOFLAV004 NAR NAR NAR) were not due to differences in growth rate but rather differences in metabolic activity. Furthermore, we confirm that even with increasing naringenin concentration in the cultures, sufficient levels to inhibit *S. albus*

growth are not being reached. The metabolic burden of producing increased levels of naringenin also does not seem to reduce the viability and growth rate of *S. albus*.

Next, we demonstrated that increasing the number of copies of the naringenin BGC in the *S. albus* UOFLAV004 genome does significantly increase naringenin yields at both 48 h and 96 hours of culture. *S. albus* UOFLAV004 NAR NAR required 96 h of culture for its yield to be significantly different from that of *S. albus* UOFLAV004 NAR. This is not surprising since past experiments in our group have demonstrated that while wild-type *S. albus* reaches its maximum naringenin yields at 48 h, the *S. albus* UOFLAV004 strain requires 96 to 120 h to reach its peak naringenin yield (data not yet published). The reason for this phenomenon has not yet been established but collecting *S. albus* UOFLAV004 culture samples every 24 h over 7 days has demonstrated a sharp uptick in naringenin yield at 96-120 h. Because of this, it is important to compare the strains at their peak yield times, which in the case of *S. albus* UOFLAV004 is 96 h. Thus, the fact that *S. albus* UOFLAV004 NAR NAR did not produce significantly more than *S. albus* UOFLAV004 NAR at 48 h is not a problem. What was surprising was that *S. albus* UOFLAV004 NAR NAR NAR produced so much more naringenin than *S. albus* UOFLAV004 NAR that the difference was already significant at 48 h of culture.

When adjusting the naringenin yields by dividing by milligrams of dry biomass, we saw the same pattern as with the unadjusted total naringenin yields. This once again confirms that the differences in naringenin production are due to differences in metabolic activity between the samples rather than differences in numbers of bacterial cells. Unfortunately, the differences in naringenin yield per mg of dry biomass between the three strains were not statistically significant due to large amount of variation between replicates. This indicates a large amount of heterogeneity within our experimental setup, something that is not uncommon when working with *S. albus*. Future experiments with more replicates will be necessary to account for this heterogeneity.

The difference in yield between *S. albus* UOFLAV004 NAR NAR and *S. albus* UOFLAV004 NAR NAR NAR is not significant at any point. This indicates the likelihood of diminishing returns from further naringenin BGC integrations. It is likely that a different bottleneck is being approached. A lack of naringenin precursors could be a limiting factor. If this is the case, then adding precursors such as tyrosine to the culture media should

significantly increase naringenin yields in the *S. albus* UOFLAV004 NAR NAR NAR strain and could also help it to maintain its peak rate of naringenin production over a longer period of time. Separately, it is possible that the *S. albus* UOFLAV004 cells are degrading the additional naringenin BGC coded enzymes due to their increased concentration within the cell or due to the depletion of tyrosine and other amino acids within the culture media (47). Measuring the concentrations of the tyrosine ammonia lyase, 4-coumaric acid-CoA ligase, chalcone synthase, and chalcone isomerase enzymes would shed light as to whether these enzymes increase in concentration with increasing number of naringenin BGC integrations.

It should be noted that the method of naringenin extraction in this experiment was different than that of the previous experiments. In previous experiments naringenin was extracted from both the bacterial cell pellet and the culture supernatant, while in this experiment only the pellet was used in the extraction. For this reason, the naringenin yields in this experiment are significantly lower than in the previous experiments and as such, comparisons between yields in this experiment and the previous experiments are not possible. Because the extraction protocol was the same for all the samples within this experiment, the samples are still comparable and the differences in naringenin yield between samples with different levels of integration are still valid. Here we see that increasing the number of BGC integrations from 1 to 3 increased naringenin yield at 96 h from 2.043 mg/L to 4.635 mg/L, which is a 2.27x increase. Increasing from 2 to 3 integration sites appears to increase naringenin yields at 96 h from 3.789 mg/L to 4.635 mg/L, which is a 1.22x increase, but this is not statistically significant.

Also, the *S. albus* UOFLAV004 NAR NAR NAR strain selection experiment (section 5.3.3) did undergo the same extraction protocol as the previous experiments and is thus comparable to them. It demonstrates that, between the metabolic modifications given to the *S. albus* UOFLAV004 strain and the use of three integration sites, we have been able to increase naringenin yields from 2.769 mg/L (*S. albus* WT NAR pSAM2) to 6.410 mg/L (*S. albus* UOFLAV004 NAR NAR NAR clone 1) at 48 h of culture time. This is a 2.3x increase, but it would likely be much greater if we compare each sample at its peak naringenin yield (96 h for *S. albus* UOFLAV004 and 48 h for *S. albus* wild type). More tests will be needed to confirm this.

7. Conclusions

- We successfully generated a modular, easily modifiable plasmid in the SEVA architectural model that contains hygromycin resistance genes for selection and the gene for pSAM2 integrase for integration. We also confirmed its integrative functionality and its use for induction of naringenin production in *S. albus*.
- We confirmed that there are differences in naringenin production depending on where in the genome the BGC is integrated. More tests are necessary to confirm why this phenomenon occurs as well as whether it is a broad occurrence or naringenin specific. More tests will also be necessary to confirm whether this information can be used to rationally design complex biosynthesis pathways and increase product yields.
- Lastly, we demonstrated that increasing the number of copies of the naringenin BGC in the *S. albus* UOFLAV004 genome significantly increases naringenin yields. Each further copy of the naringenin BGC induces a diminishing return in the rate of production, indicating that other bottlenecks to naringenin production are being met. Further tests are needed to identify the nature of these bottlenecks (such as precursors depletion) and see if mixing the multiple integration strategy with other strategies can cause further improvements in naringenin yield.
- Finally, we demonstrate that between the modifications to the *S. albus* UOFLAV004 strain (data not yet published and not included in this work) and the use of multiple integration sites, we have achieved a 2.3x increase in naringenin yield compared with the wild type *S. albus* strain with only one integrated naringenin BGC at 48 h of culture time.

8.References

1. Mathesius U. Flavonoid Functions in Plants and Their Interactions with Other Organisms. Plants. 2018 Jun;7(2):30.

2. Panche AN, Diwan AD, Chandra SR. Flavonoids: an overview. J Nutr Sci [Internet]. 2016 ed [cited 2022 Jun 1];5. Available from:

https://www.cambridge.org/core/journals/journal-of-nutritional-science/article/flavonoids-an-overview/C0E91D3851345CEF4746B10406908F52

3. Kim Y, Je Y. Flavonoid intake and mortality from cardiovascular disease and all causes: A meta-analysis of prospective cohort studies. Clin Nutr ESPEN. 2017 Aug 1;20:68–77.

4. Chang H, Lei L, Zhou Y, Ye F, Zhao G. Dietary Flavonoids and the Risk of Colorectal Cancer: An Updated Meta-Analysis of Epidemiological Studies. Nutrients. 2018 Jul;10(7):950.

5. Maher P. The Potential of Flavonoids for the Treatment of Neurodegenerative Diseases. Int J Mol Sci. 2019 Jan;20(12):3056.

6. Darband SG, Kaviani M, Yousefi B, Sadighparvar S, Pakdel FG, Attari JA, et al. Quercetin: A functional dietary flavonoid with potential chemo-preventive properties in colorectal cancer. J Cell Physiol. 2018;233(9):6544–60.

7. Devi KP, Rajavel T, Nabavi SF, Setzer WN, Ahmadi A, Mansouri K, et al. Hesperidin: A promising anticancer agent from nature. Ind Crops Prod. 2015 Dec 15;76:582–9.

8. Fernández J, Silván B, Entrialgo-Cadierno R, Villar CJ, Capasso R, Uranga JA, et al. Antiproliferative and palliative activity of flavonoids in colorectal cancer. Biomed Pharmacother. 2021 Nov 1;143:112241.

9. Nishiumi S, Miyamoto S, Kawabata K, Ohnishi K, Mukai R, Murakami A, et al. Dietary flavonoids as cancer-preventive and therapeutic biofactors. Front Biosci-Sch. 2011 Jun 1;3(4):1332–62.

10. Fernández J, García L, Monte J, Villar CJ, Lombó F. Functional Anthocyanin-Rich Sausages Diminish Colorectal Cancer in an Animal Model and Reduce Pro-Inflammatory Bacteria in the Intestinal Microbiota. Genes. 2018 Mar;9(3):133.

11. Stracy M, Snitser O, Yelin I, Amer Y, Parizade M, Katz R, et al. Minimizing treatment-induced emergence of antibiotic resistance in bacterial infections. Science. 2022 Feb 25;375(6583):889–94.

12. Sonnenburg ED, Sonnenburg JL. The ancestral and industrialized gut microbiota and implications for human health. Nat Rev Microbiol. 2019 Jun;17(6):383–90.

13. Hviid A, Svanström H, Frisch M. Antibiotic use and inflammatory bowel diseases in childhood. Gut. 2011 Jan 1;60(1):49–54.

14. Górniak I, Bartoszewski R, Króliczewski J. Comprehensive review of antimicrobial activities of plant flavonoids. Phytochem Rev. 2019 Feb 1;18(1):241–72.

15. Selma MV, Espín JC, Tomás-Barberán FA. Interaction between Phenolics and Gut Microbiota: Role in Human Health. J Agric Food Chem. 2009 Aug 12;57(15):6485–501.

16. Harborne JB, Williams CA. Advances in flavonoid research since 1992. Phytochemistry. 2000 Nov 1;55(6):481–504.

17. Wen L, Jiang Y, Yang J, Zhao Y, Tian M, Yang B. Structure, bioactivity, and synthesis of methylated flavonoids. Ann N Y Acad Sci. 2017;1398(1):120–9.

18. Marín L, Gutiérrez-del-Río I, Villar CJ, Lombó F. De novo biosynthesis of garbanzol and fustin in Streptomyces albus based on a potential flavanone 3-hydroxylase with 2-hydroxylase side activity. Microb Biotechnol. 2021;14(5):2009–24.

19. Marín L, Gutiérrez-del-Río I, Entrialgo-Cadierno R, Villar CJ, Lombó F. De novo biosynthesis of myricetin, kaempferol and quercetin in Streptomyces albus and Streptomyces coelicolor. PLOS ONE. 2018 Nov 15;13(11):e0207278.

20. Marín L, Gutiérrez-del-Río I, Yagüe P, Manteca Á, Villar CJ, Lombó F. De Novo Biosynthesis of Apigenin, Luteolin, and Eriodictyol in the Actinomycete Streptomyces albus and Production Improvement by Feeding and Spore Conditioning. Front Microbiol [Internet]. 2017 [cited 2022 Jun 3];8. Available from: https://www.frontiersin.org/article/10.3389/fmicb.2017.00921

21. Teixeira A, Eiras-Dias J, Castellarin SD, Gerós H. Berry Phenolics of Grapevine under Challenging Environments. Int J Mol Sci. 2013 Sep;14(9):18711–39.

22. Nabavi SM, Šamec D, Tomczyk M, Milella L, Russo D, Habtemariam S, et al. Flavonoid biosynthetic pathways in plants: Versatile targets for metabolic engineering. Biotechnol Adv. 2020 Jan 1;38:107316.

23. Koopman F, Beekwilder J, Crimi B, van Houwelingen A, Hall RD, Bosch D, et al. De novo production of the flavonoid naringenin in engineered Saccharomyces cerevisiae. Microb Cell Factories. 2012 Dec 8;11(1):155.

24. Salehi B, Fokou PVT, Sharifi-Rad M, Zucca P, Pezzani R, Martins N, et al. The Therapeutic Potential of Naringenin: A Review of Clinical Trials. Pharmaceuticals. 2019 Mar;12(1):11.

25. Zaburannyi N, Rabyk M, Ostash B, Fedorenko V, Luzhetskyy A. Insights into naturally minimised Streptomyces albus J1074 genome. BMC Genomics. 2014 Feb 5;15(1):97.

26. Manteca A, Sanchez J, Jung HR, Schwämmle V, Jensen ON. Quantitative Proteomics Analysis of Streptomyces coelicolor Development Demonstrates That Onset of Secondary Metabolism Coincides with Hypha Differentiation *. Mol Cell Proteomics. 2010 Jul 1;9(7):1423–36.

27. Che S, Men Y. Synthetic microbial consortia for biosynthesis and biodegradation: promises and challenges. J Ind Microbiol Biotechnol. 2019 Oct 1;46(9–10):1343–58.

28. Li L, Wei K, Liu X, Wu Y, Zheng G, Chen S, et al. aMSGE: advanced multiplex site-specific genome engineering with orthogonal modular recombinases in actinomycetes. Metab Eng. 2019 Mar 1;52:153–67.

29. Merrick CA, Zhao J, Rosser SJ. Serine Integrases: Advancing Synthetic Biology. ACS Synth Biol. 2018 Feb 16;7(2):299–310.

30. Groth AC, Calos MP. Phage Integrases: Biology and Applications. J Mol Biol. 2004 Jan 16;335(3):667–78.

31. The integrated conjugative plasmid pSAM2 of Streptomyces ambofaciens is related to temperate bacteriophages. EMBO J. 1989 Mar;8(3):973–80.

32. Aubry C, Pernodet JL, Lautru S. Modular and Integrative Vectors for Synthetic Biology Applications in Streptomyces spp. Appl Environ Microbiol. 2019 Aug 15;85(16):e00485-19.

33. Jeong KS, Ahn J, Khodursky AB. Spatial patterns of transcriptional activity in the chromosome of Escherichia coli. Genome Biol. 2004 Oct 27;5(11):R86.

34. PARDO JM, MALPARTIDA F, RICO M, JIMÉNEZ A 1985. Biochemical Basis of Resistance to Hygromycin B in Streptomyces hygroscopicus – the Producing Organism. Microbiology. 131(6):1289–98.

35. DW Russell, J Sambrook. Molecular cloning: a laboratory manual. Fourth Edition. CSHL Press; 2001.

36. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009 May;6(5):343–5.

37. Iverson SV, Haddock TL, Beal J, Densmore DM. CIDAR MoClo: Improved MoClo Assembly Standard and New E. coli Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology. ACS Synth Biol. 2016 Jan 15;5(1):99–103.

38. Green R, Rogers EJ, Lorsch J. Chapter Twenty Eight - Transformation of Chemically Compentent E. coli. In: Methods in Enzymology. Academic Press; 2013.

39. T. Kieser, M.J. Bibb, M.J. Buttner, K.F. Chater & D.A. Hopwood. PRACTICAL STREPTOMYCES GENETICS. Norwich, UK: The John Innes Foundation; 2000.

40. Myronovskyi M, Tokovenko B, Brötz E, Rückert C, Kalinowski J, Luzhetskyy A. Genome rearrangements of Streptomyces albus J1074 lead to the carotenoid gene cluster activation. Appl Microbiol Biotechnol. 2014 Jan 1;98(2):795–806.

41. Bilyk B, Luzhetskyy A. Unusual site-specific DNA integration into the highly active pseudo-attB of the Streptomyces albus J1074 genome. Appl Microbiol Biotechnol. 2014 Jun 1;98(11):5095–104.

42. Rokytskyy I, Koshla O, Fedorenko V, Ostash B. Decoding options and accuracy of translation of developmentally regulated UUA codon in Streptomyces: bioinformatic analysis. SpringerPlus. 2016 Jul 4;5(1):982.

43. Baltz RH. Streptomyces temperate bacteriophage integration systems for stable genetic engineering of actinomycetes (and other organisms). J Ind Microbiol Biotechnol. 2012 May 1;39(5):661–72.

44. Chater KF, Wilde LC. Restriction of a bacteriophage of Streptomyces albus G involving endonuclease SalI. J Bacteriol. 1976 Nov;128(2):644–50.

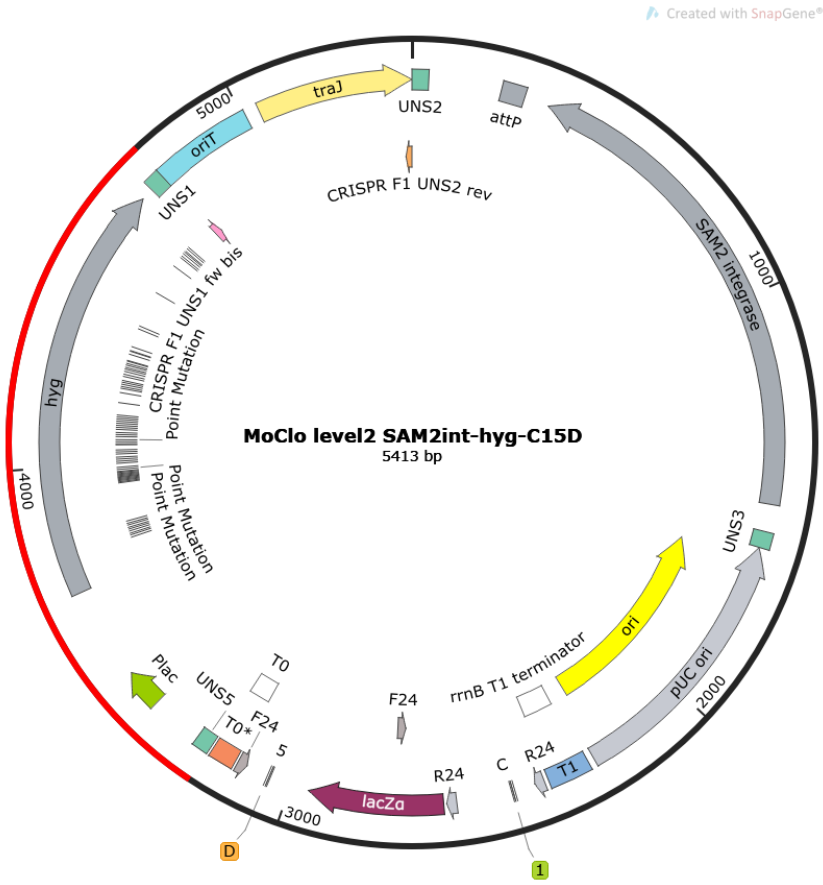
45. González A, Rodríguez M, Braña AF, Méndez C, Salas JA, Olano C. New insights into paulomycin biosynthesis pathway in Streptomyces albus J1074 and generation of novel derivatives by combinatorial biosynthesis. Microb Cell Factories. 2016 Mar 21;15(1):56.

46. D3veloperSCS_SEVA. Home [Internet]. SEVA plasmids - Standard European Vector Architecture. [cited 2022 Jun 13]. Available from: http://seva-plasmids.com/

47. Bradley MO. Regulation of protein degradation in normal and transformed human cells. Effects of growth state, medium composition, and viral transformation. J Biol Chem. 1977 Aug;252(15):5310–5.

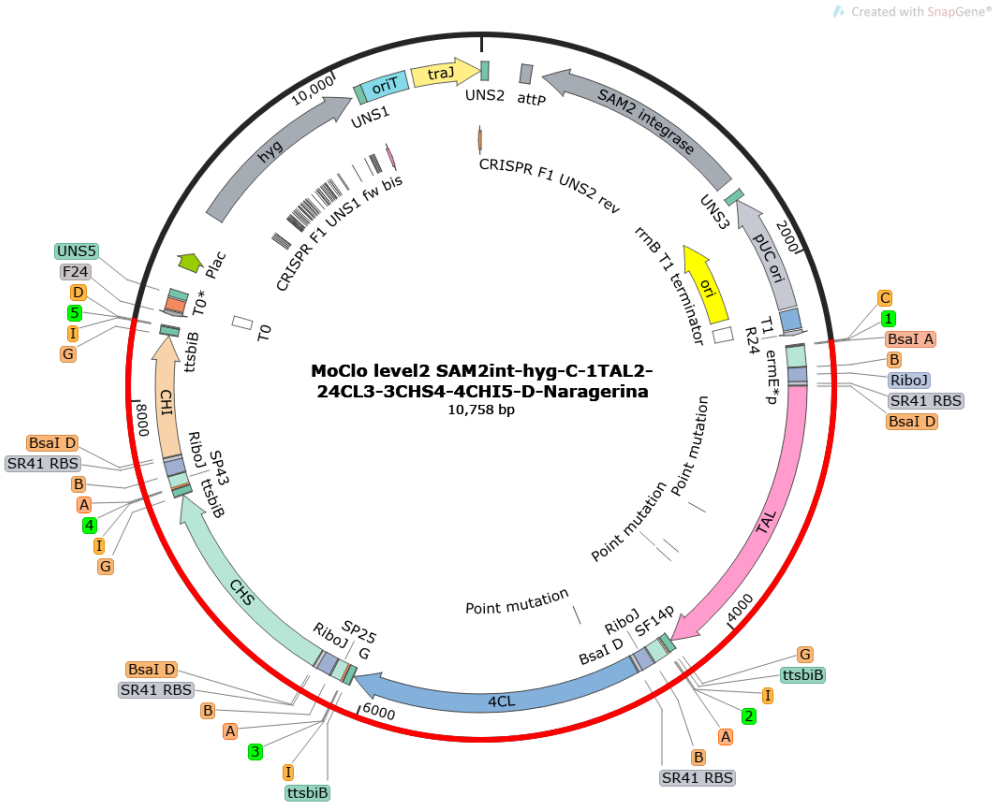
9. Supplementary Data

Supplementary Figure 1

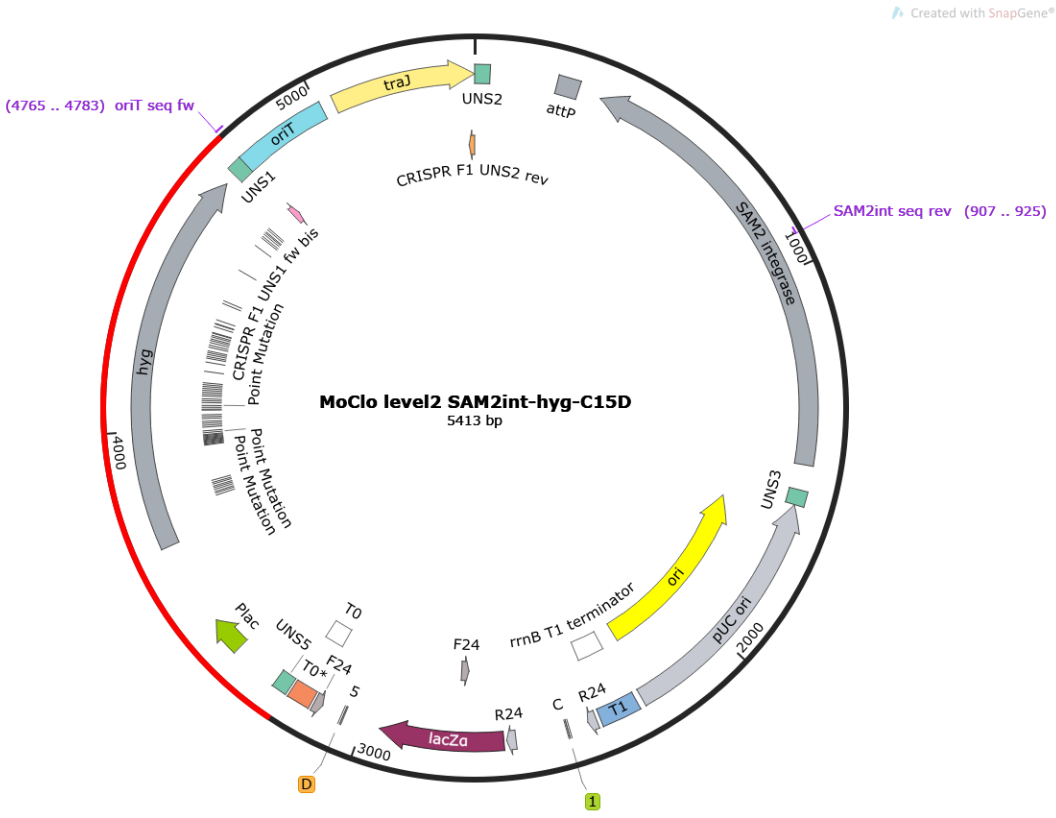


Supplementary Figure 2.

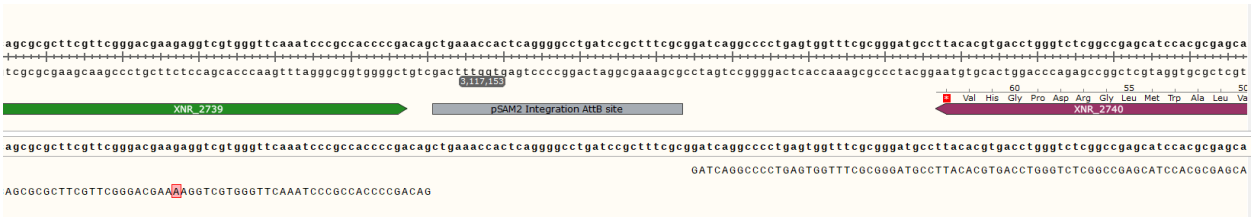
B



Supplementary Figure 3



Supplementary Figure 4



10. Acknowledgements

First, a tremendous thank you to my family for all their love and support. To my mother for helping me make the jump to and get settled in Spain. To my dads for supporting me in my decision to leave home to pursue graduate school abroad. I love you all and miss you guys.

Second, to everyone in the master's administration that helped make this year a tremendous learning experience. Especially to Xose and Ana, for going above and beyond in helping me through the application and enrollment process. I would not be in this program if it were not for your efforts.

Third, to Felipe, Suhui, and all the other BIONUC lab members. Thank you for being the best research group I could have asked for. In every moment I have felt accepted in the group and supported in my TFM work. The positive and energetic atmosphere within this group is truly infectious and a joy to be a part of. I can't wait to start on my PhD with you and be a part of the team for many more years.

Fourth, to my friends in the master's program. I was nervous as to how easily I would be able to make friends amongst a group of people that were a few years younger than me and from a different continent. Within two days I realized that I had nothing to fear. Thank you all for including me in your lives and being truly amazing friends. It is thanks to you that this year has been nonstop fun and that I am so excited to remain in Oviedo for multiple more years.

To all my other friends in Oviedo, the same goes to you.

Fifth, to my old mentors Josh, John, Lichao, and everyone else in Stanford Otolaryngology/CZ Biohub mass spec platform (past and present). Thank you for all your efforts in developing me as a scientist during the years I was with you. It is thanks to you that I fell in love with science and decided to pursue graduate school. Also, thank you for allowing me to join the occasional lab meeting since I have left and for your continuing to support me in my academic career.

Finally, to all my friends from back home in California. Thank you all for being so supportive of my leaving to pursue graduate school abroad. I miss you all and can't wait to see you in September. I hope some of you will be able to take advantage of my being in Oviedo for multiple more years and come visit. Thank you for always keeping in touch, every time we talk it seems as if time doesn't pass.

Once again, to everyone that has supported me up until now, no matter how large or small our interaction, thank you. This work is devoted to all of you.

11. Author's Declaration of Originality

I hereby certify that I am the sole author of this TFM. I certify that that all the data and discussion included in this work is original unless stated otherwise and properly sited. To the best of my knowledge, my TFM does not infringe upon anyone's copyright nor violate any proprietary rights and any ideas, techniques, quotations, or any other material from the work of other people included in my TFM, published or otherwise, are fully acknowledged in accordance with the standard referencing practices.