

Supplementary Material

Two actinobacterial G protein systems control multicellular mycelial growth and morphology in *Streptomyces venezuelae*

Dima Massri, Rikard Osgyan, Veronica Ekdahl, and Klas Flärdh

Content

Supplementary Methods

Whole-genome sequence analysis

Transcriptomic data analysis

Supplementary Results

Whole-genome sequencing analysis of the *cvn*-free strains revealed an undesired 62 kb deletion

Supplementary Figures

Figures S1-S16

Supplementary Tables

Tables S1-S3

Supplementary Data Files (separate files)

Tables S4-S14

Supplementary References

Supplementary Methods

Whole-genome sequence analysis

Paired-end Illumina reads were aligned to *S. venezuelae* wild-type NRRL B-65442 reference genome, available in the NCBI database under accession number NZ_CP018074.1. Single-nucleotide variants were called using Snippy v4.6.0 ¹, which internally aligns reads using BWA-MEM. The following thresholds were applied for detection: minimum coverage 10, minimum alternative allele fraction 0.9, minimum base quality 100, and minimum mapping quality 50. The coordinate-sorted BAM files produced by Snippy/ BWA-MEM were subsequently used for coverage analysis and visualization. Read-mapping quality was computed with Qualimap v2.2.2 ².

For structural variant (SV) detection, two complementary tools, LUMPY v0.2.13 ³ and DELLY v1.3.3 ⁴ were independently used to reduce method-specific bias. Several preprocessing steps were done before SV detection. Paired-end sequencing reads were first aligned to the reference genome using the BWA-MEM algorithm in BWA v0.7.17 ⁵ with default parameters. Alignment files were processed using SAMtools v1.20 ⁶ to generate BAM files. Duplicate reads were identified using Samblaster v0.1.26 ⁷, which marked reads originating from the same DNA fragment. Two datasets were generated for each sample: (i) Duplicate-marked dataset, where duplicates were kept but marked and (ii) Duplicate-removed dataset, where duplicates were excluded. All BAM files were sorted both by read name and genomic coordinate using SAMtools v1.20. Downstream analysis and final SV tables were obtained using the duplicate-marked dataset to preserve SV supporting evidence. The duplicate-removed dataset was only used to extract supporting reads of SV evidence. Two types of supporting reads were extracted: (i) Discordant read pairs, defined as paired reads with unexpected orientation or insert size and (ii) Split reads, defined as individual reads aligning to two genomic locations, indicating a breakpoint. Discordant reads were extracted using SAMtools filtering, whereas split reads were obtained using `extractSplitReads_BwaMem` included with LUMPY. For each sample, insert size distributions were estimated empirically using `pairend_distro.py` also included with LUMPY. Mean insert size and standard deviation derived from these distributions were used as parameters during structural variant calling.

Upon completion of preprocessing steps, LUMPY and DELLY were executed for SV detection. LUMPY integrates paired-end and split-read evidence probabilistically to infer breakpoint locations. For LUMPY, structural variants were called using both discordant and split-read BAM files with a minimum mapping quality threshold of 20. DELLY was run using coordinate-sorted BAM files with a minimum mapping quality threshold of 20. The software jointly evaluates paired-end and split-read signals to detect deletions and other structural rearrangements. Raw variant calls were converted to sorted and indexed VCF files using BCFtools v1.21 ⁶ and variants were filtered based on read support: (i) LUMPY variants with fewer than 10 total supporting reads were labeled as low quality and (ii) DELLY variants supported by less than 10 combined paired-end and split reads were labeled as low quality. Only variants passing these filters were retained for downstream analysis. Variants that passed filtering were merged across samples using SURVIVOR v1.0.7 ⁸ allowing breakpoint differences within a 500 bp distance threshold. High-confidence variants were further manually inspected in IGV (Integrative Genomics Viewer v2.19.7) ^{9,10} retaining only events that showed the expected signatures, for example clusters of large-insert discordant pairs for deletions, mirrored same-orientation (RR/FF) pair clusters at both breakpoints for inversions, and reciprocal RL/LR clusters at both breakpoints accompanied by a read-depth increase across the interval for duplications. Calls that lacked these patterns or showed only local boundary artefacts were removed. Additionally, SV candidates larger than 100 kb were considered artefacts without manual inspection.

For visualization of genome coverage and genomic variants, aligned reads in BAM format were converted to genome-wide coverage tracks using the `genomecov` function of BEDTools v2.31.1¹¹ with the `-bga` and `-split` options, yielding BedGraph files that include zero-coverage intervals and respect split alignments. The resulting BedGraph files were processed using Pandas (v3.0.0)¹² and NumPy v2.4.1¹³ python packages to obtain coverage values for genomic intervals (Python v3.14.2)¹⁴. Coverage values were normalized by dividing each interval by the maximum coverage in the corresponding sample. Coverage tracks were visualized with matplotlib (v3.10.8) in both linear and circular formats. Linear plots showed normalized genome-wide coverage, with ≥ 50 bp zero-coverage regions automatically highlighted. Circular plots used polar coordinates to display multiple samples as concentric radial tracks, with highlighted ≥ 50 bp zero-coverage regions and genomic coordinate markers along the outer axis. In the plots where SNPs and SVs were included, these annotations were mapped to genomic coordinates and visualized as layered tracks.

Transcriptomic data analysis

For RNAseq data analysis, FASTQ files containing raw reads were initially processed to remove adapters and low-quality bases using TrimGalore v0.6.10¹⁵. TrimGalore was run with default parameters, Phred quality threshold of 20 and reads shorter than 20 bp were discarded. Trimmed reads were filtered to remove tRNA and rRNA sequences that persisted despite the rRNA depletion process. This was done by first using BEDTools v2.31.1 to build a custom reference genome database combining all rRNA and tRNA sequences extracted from *S. venezuelae* genome annotation file (gff3) corresponding to the genome archived in NCBI under accession number NZ_CP018074.1. Trimmed reads were then mapped to the rRNA-tRNA custom genome using Bowtie2 v2.5.4¹⁶ with sensitive settings (`--sensitive`) and the unaligned pairs, which are rRNA/tRNA-depleted and enriched for mRNA sequences, were saved as FASTQ files. Trimmed rRNA/tRNA-depleted reads were further down-sampled to equal numbers per sample using `reformat.sh` from the BBDMap/BBTools suite¹⁷. The minimum number of cleaned reads across all samples and both mates was determined to decide on a down-sampling target and reads were randomly subsampled from each FASTQ file to normalize sequencing depth across samples, keeping the overall expression distribution intact for downstream mapping and counting. FastQC v0.12.1¹⁸ was run on raw reads and after each processing step (Trimming, rRNA/tRNA removal, and down-sampling), and read counts were recorded at each step to monitor data quality throughout the workflow. The final processed reads were mapped to reference genome using Bowtie2 with sensitive settings (`--sensitive`), and post-alignment quality check was done using SAMtools `flagstat` and `stats` subcommands of SAMtools v1.22.1.

Sorted alignment BAM files were used to generate gene-level read counts using FeatureCounts from the Subread package v2.1.1¹⁹. FeatureCounts was run specifying paired-end reads (`-p`), reverse-stranded libraries (`-s 2`), and counting CDS features using `locus_tag` as the gene identifier. Filtering for low-count genes was performed to keep only genes with ≥ 10 counts in ≥ 2 samples to reduce noise from very low-expressed genes that can introduce variability in downstream differential expression analysis. All quality control and count metrics were then aggregated using MultiQC v1.32²⁰.

Exploratory analysis was later done on count data to assess variance, potential batch effect, outliers, and relationship between the samples. FeatureCounts-generated filtered counts were initially transformed to VST-normalized counts (Variance stabilizing transformation) using the DESeq2 package from Bioconductor v1.38.0²¹ to and make data suitable for exploratory visualization. Assessment of the VST-normalized data revealed one outlier per condition, each corresponding to one of the two sequencing batches. Therefore, batch correction for visualization was performed using `limma`'s `removeBatchEffect` from Bioconductor v3.54.0²², which fits a linear model to estimate and subtract

batch-associated variation for each gene, yielding batch-adjusted variance-stabilized expression values. Batch-corrected VST-normalized counts were used to generate PCA plots to visualize similarity between samples based on gene expression variance. PCA visualization was performed in R v4.2.3²³ using limma's *removeBatchEffect* for batch adjustment and ggplot2 v3.5.1²⁴, data.table v1.15.4²⁵, and reshape2 v1.4.4²⁶ R packages for data handling and plotting.

Differential expression analysis was performed using DESeq2 from Bioconductor v1.38.0 with wild type set as the reference and the mutant compared against it. FeatureCounts-generated filtered counts were used as input and DESeq2 was run with batch effect handled through the design formula ($\sim$ batch + condition) to model batch as a covariate without altering count data. Gene product annotation was added to DESeq2 results by parsing *S. venezuelae* genome annotation file (gff3) to extract CDS features and their associated "product" fields and merging annotation with DESeq2 output tables. Filtered differentially expressed (DE) genes result tables (Supplementary Tables S6 and S7) were generated by filtering annotated DESeq2 output tables based on log2 fold-change and adjusted *p* value cutoffs. Two annotated DE gene tables were generated: one filtered using permissive thresholds (log2 fold change ≥ 1 with adjusted *p* value ≤ 0.05), and a second filtered using stricter thresholds (log2 fold change ≥ 2 with adjusted *p* value ≤ 0.01). Additional columns for direction of change (up or down) and a "significance" column were appended to these tables. These data-processing steps were performed in R using the data.table v1.15.4 and dplyr v1.1.4²⁷ packages. For visualization of global expression patterns, volcano plots were generated in R using data.table v1.15.4 for data handling and ggplot2 v3.5.1 for visualization, applying the same permissive and strict thresholds of log2 fold change and adjusted *p* value described above. Genes corresponding to the deleted *cvn* loci in the mutant strain were omitted from these plots.

For functional interpretation, protein sequences of all expressed genes (i.e., genes present in FeatureCounts results and retained after filtering) were extracted from the reference protein FASTA using the Python packages (Python v3.9.23) Pandas v3.0.1 for processing CSV files, and SeqKit v2.13.0²⁸ for sequence retrieval. Functional annotation of all expressed proteins was performed using eggNOG-mapper (emapper.py) v2.1.13^{29,30} with the DIAMOND search mode, which uses fast protein sequence alignment to assign orthology and functional categories. EggNOG databases were downloaded from the official repository (<http://eggno5.embl.de>) and annotations were generated for all expressed proteins. The resulting full EggNOG output was then used to extract functional annotations for all DESeq2-identified DE genes and for filtered DE gene subsets defined by the permissive and strict thresholds described above. Statistics and parameters (log2 fold change, adjusted *p* value, and direction of regulation) were merged with EggNOG-derived functional information, and DE genes were subsequently divided into up-regulated and down-regulated genes. Counts of COG categories, KEGG pathways, and GO terms were generated for all DE genes and for up-regulated and down-regulated subsets using the Python packages Pandas v3.0.1 and Python's built-in collections.Counter (Python v3.9.23). Genes belonging to specific functional categories (COG, KEGG, or GO) were then extracted from these annotated datasets using Pandas.

Because COG provided the most informative functional signal, counts of COG terms generated in the previous step were used for functional enrichment analysis, in which each of the up-regulated and down-regulated gene subset (filtered by the log2 fold change and adjusted *p* value thresholds described above) was compared against the annotated background using Fisher's exact test with FDR correction. Compound COG terms were split into single-letter categories prior to testing, and all calculations were done in Python using Pandas v3.0.1, SciPy v1.17.1³¹, and statsmodels v0.14.6³².

Enrichment results were visualized as bubble plots after computing enrichment metrics (enrichment ratio, $-\log_{10}[\text{FDR}]$, and a significance flag for $\text{FDR} < 0.1$), retaining only terms represented by at least three genes. Data preparation for plotting was done in Python v3.9.23, using Pandas v3.0.1 and NumPy v2.4.2. Bubble plots displayed enrichment ratio (x-axis) calculated as $(\text{gene set count} / \text{total genes in set}) \div (\text{background count} / \text{total background genes})$, COG category (y-axis), gene count (bubble size), $-\log_{10}[\text{FDR}]$ (color), and were faceted by direction of regulation. Gene-level dot plots were produced for selected genes, using DESeq2 outputs with computed $-\log_{10}(\text{padj})$, where genes were shown on the x-axis, $\log_2$ fold change on the y-axis, and baseMean and $-\log_{10}(\text{padj})$ encoded as bubble size and color, respectively. Both bubble plots and dot plots were generated in R v4.2.3 using ggplot2 v3.5.2 package within the tidyverse v.0.4.6 ecosystem³³.

Supplementary Results

Whole-genome sequencing analysis of the *cvn*-free strains revealed an undesired 62 kb deletion

Whole-genome sequencing (WGS) analysis of the *cvn*-free strains LUV314 and LUV315 revealed the deletion of a genetic element of approximately 62 kb in size in both biological replicates (marked with an asterisk * in Fig. 1C and Supplementary Fig. S1). The deletion borders were first estimated from WGS data and further validated by performing PCR with primers flanking the predicted deletion junction and Sanger sequencing of the amplicon. Sequence alignment revealed that the element encompasses the region from 2280558 to 2342628 bp, carries a predicted site-specific integrase (*vnz10120*) and is flanked by two direct repeats that are putative target sites for the site-specific recombinase, with one of these repeats overlapping with the tRNA gene *vnz10615* in the wild-type *S. venezuelae* strain (Supplementary Fig. S2A). The undesired deletion was therefore estimated to be 62010 bp and the region encompasses 98 annotated coding sequences, 69 of which are specific to *S. venezuelae* in the StepDB database (<https://strepdb.streptomyces.org.uk>), and most of which (88 genes) encode "hypothetical proteins", lacking experimental validation of functional roles (Supplementary Table S5). Based on the presence of phage-related genes, the element is referred to as a putative prophage-like element. Since the *cvn*-free strains were constructed by sequential mutagenesis, intermediate strains were screened for the presence of the undesired deletion with two diagnostic PCRs: a deletion-specific PCR using a primer pair flanking the deletion region and producing a product only when the deletion is present, and a wild type-specific PCR with a primer inside the deleted region giving amplification only from the undeleted version. These assays showed that the deletion first appeared in $\Delta cvn1$ -2-3-9-5-7 strain and was therefore present in all downstream strains ($\Delta cvn1$ -2-3-9-5-7-4 and $\Delta cvn1$ -2-3-9-5-7-4-6) (Supplementary Fig. S2B).

Supplementary Figures

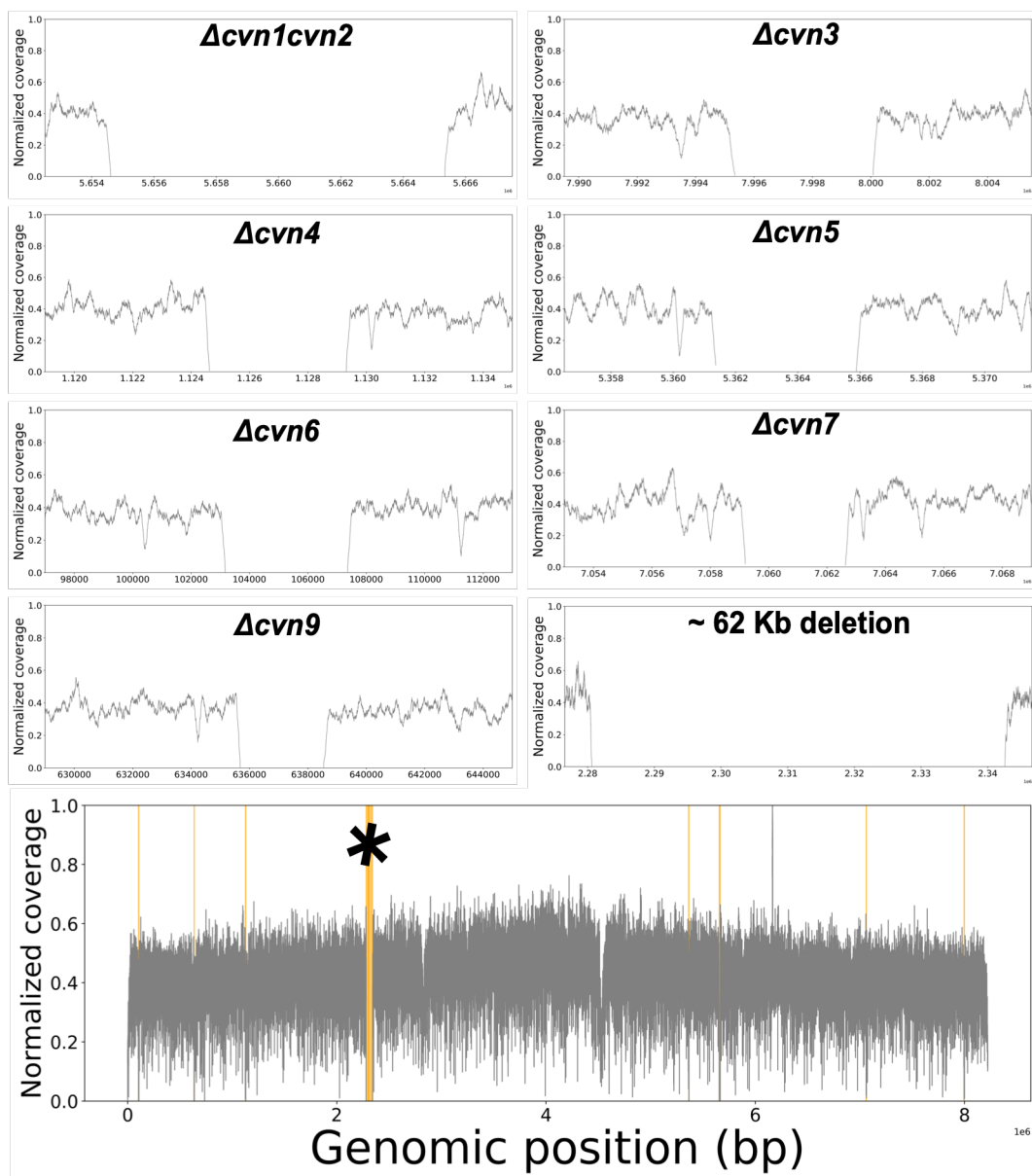


Figure S1. Systematic mutagenesis of all 8 *cvn* loci of *S. venezuelae*. Coverage plots of sequencing reads of one of the *cvn*-free mutants, strain LUV314, mapped to *S. venezuelae* NRRL B-65442 wild-type reference genome. The main line represents the genome-wide coverage (bottom), with vertical yellow lines indicating deletions and the asterisk marking the undesired 62 kb deletion. The plots above the main line show the coverage across 16 kb regions encompassing the regions of the conserved deletions and the coverage across 70 kb region encompassing the 62 kb deletion representing the excised mobile element.

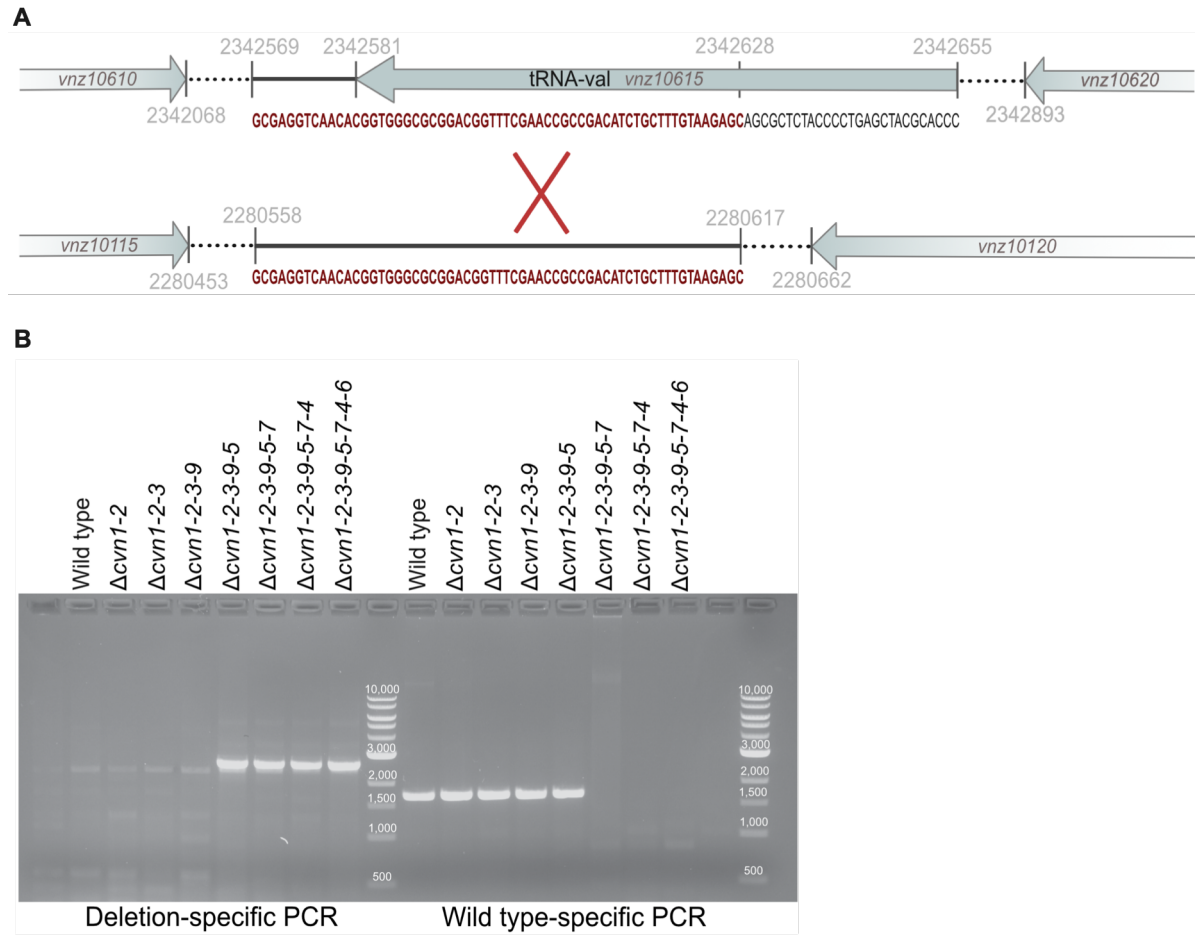


Figure S2. Whole-genome sequence analysis of the *cvn*-free strains revealed an undesired 62 kb deletion. (A) An illustration depicting the genomic position and the proposed recombination underlying the deletion. Direct repeats flanking the deletion region are shown in red and the recombination event is denoted by (X) at the recombination site. (B) Gel image displaying PCR results for detection of the deletion in the parent strains of the *cvn*-free strain. Deletion-specific PCR was done using primers flanking the deletion (primer pair KF2167/2172, product size = 2558 bp). Wild type-specific PCR was done using one primer upstream the deletion and one primer inside the wild-type sequence (primer pair KF2168/2175, product size = 1684 bp). PCR products are shown on a 1% agarose gel. A 1 kb DNA ladder was used as a molecular size marker.

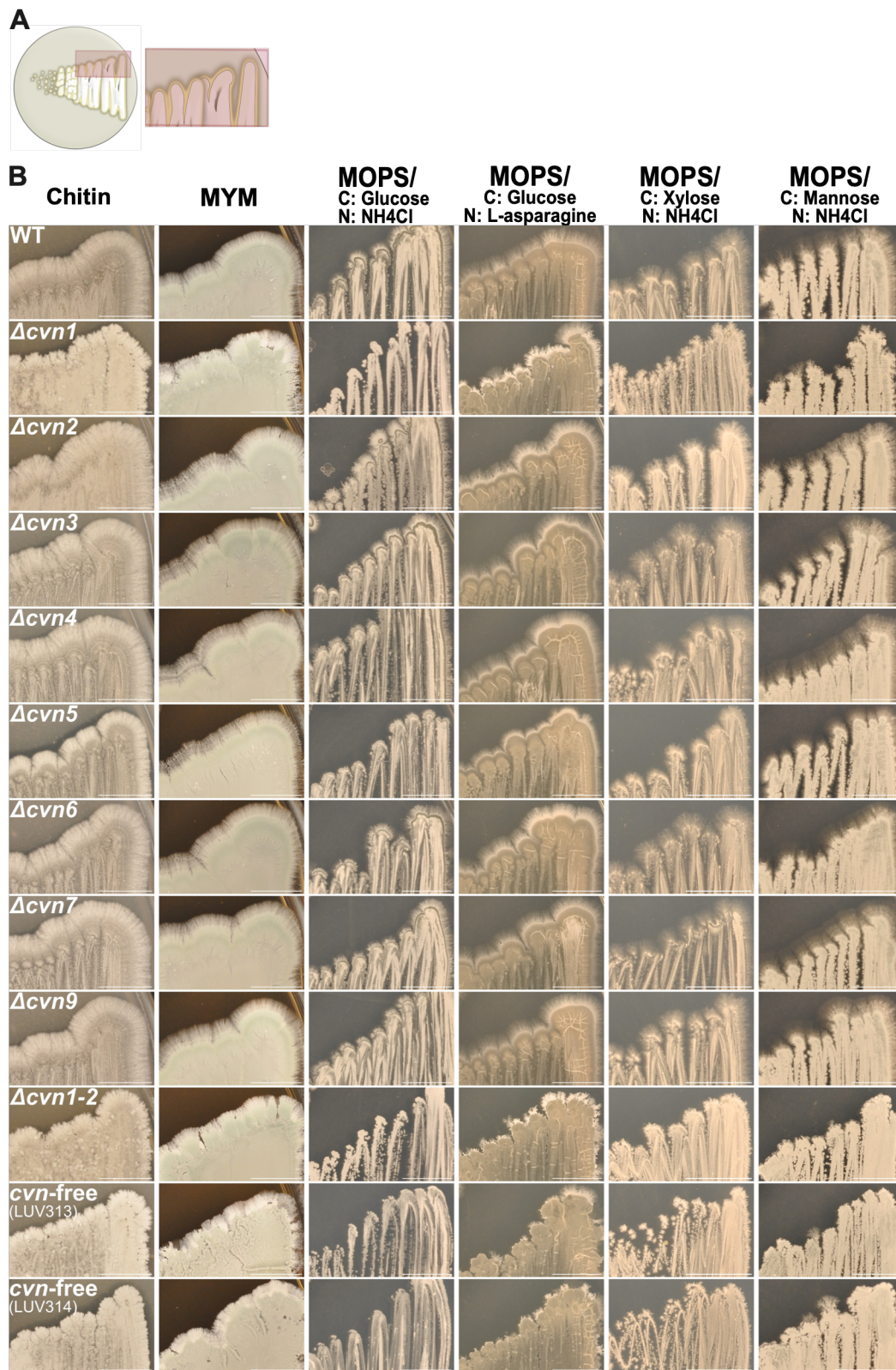


Figure S3. Screening for macroscopic phenotypes of the *cvn* mutants. (A) A sketch illustrating the process of patching on different growth media and the cropping of the edges where distinct phenotypic characteristics were photographed. (B) Images showing the cropped edges of patches of the strains, from top to bottom, *S. venezuelae* wild type, LUV258 ($\Delta cvn1$), LUV309 ($\Delta cvn2$), LUV259 ($\Delta cvn3$), LUV310 ($\Delta cvn4$), LUV276 ($\Delta cvn5$), LUV311 ($\Delta cvn6$), LUV260 ($\Delta cvn7$), LUV263 ($\Delta cvn9$), LUV256 ($\Delta(cvn1-cvn2)$), and the *cvn*-free mutant strains LUV313 and LUV314. Loopfuls of spore suspensions

of similar concentrations were patch-streaked and grown on, from left to right, chitin for 9 days, MYM for 10 days, MOPS with glucose and ammonium chloride for 9 days, MOPS with glucose and L-asparagine for 9 days, MOPS with xylose and ammonium chloride for 4 days, and MOPS with mannose and ammonium chloride for 7 days. Scale bars, 1 cm.

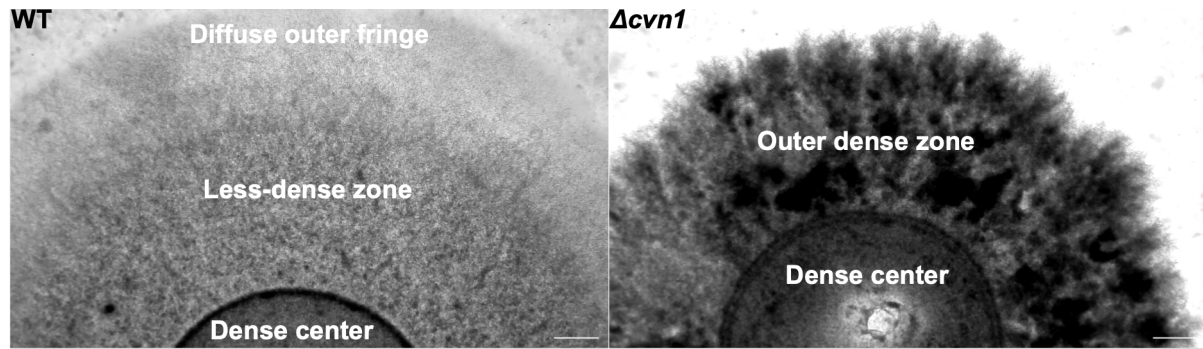


Figure S4. Macroscopic phenotype of *cvn1* deletion mutant on chitin agar. Light microscopy images showing the colony appearance with distinct concentric rings and the fringe-center distinction in wild-type patches, when grown on chitin agar (left), and the distinct colony phenotype of LUV258 ($\Delta cvn1$), with more evenly dense colonies. 5 μ l of spore suspensions of similar concentrations was spotted on agar and images were taken using a stereomicroscope system with transmitted light LED illumination after 9 days of incubation. Scale bars, 0.1 cm.

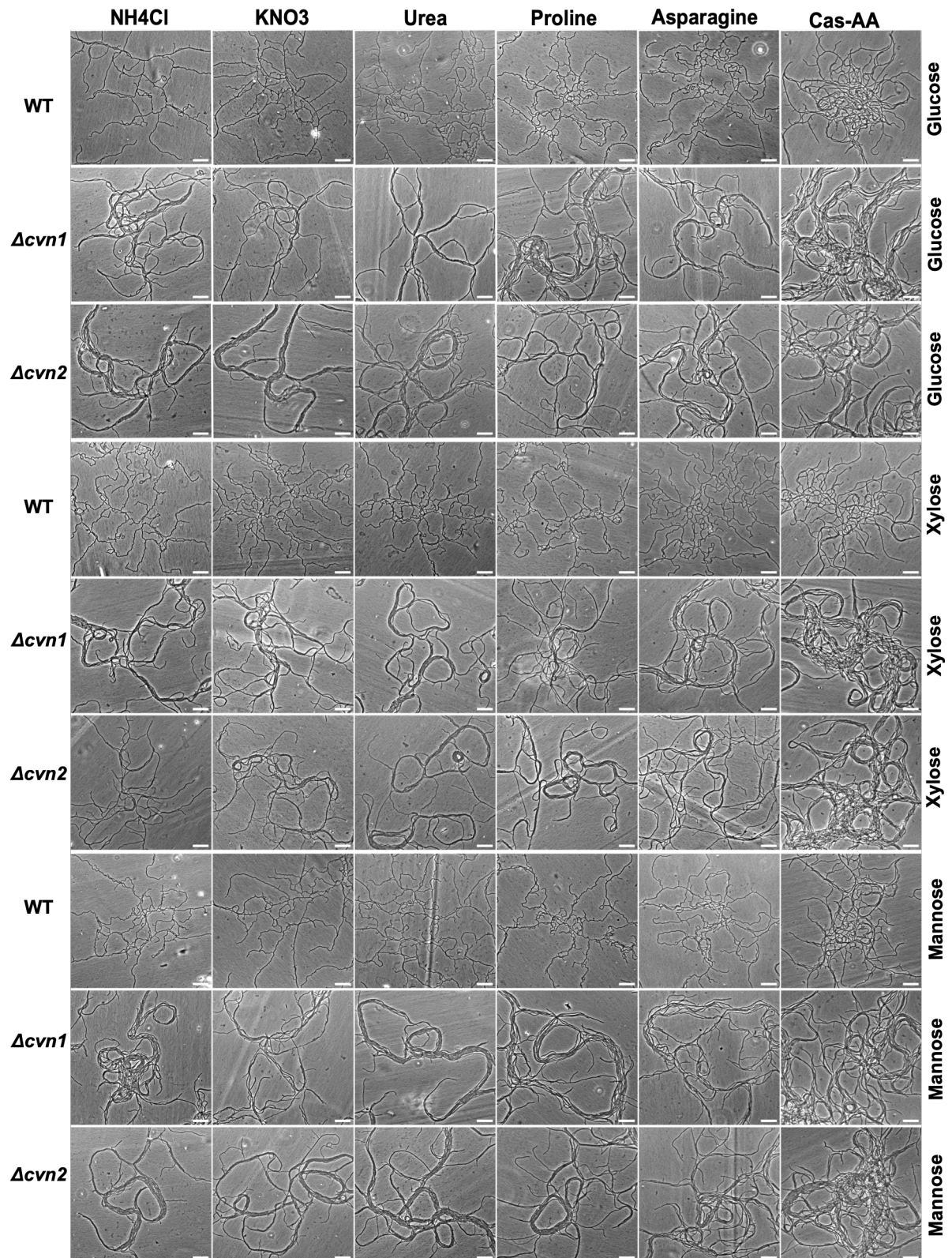
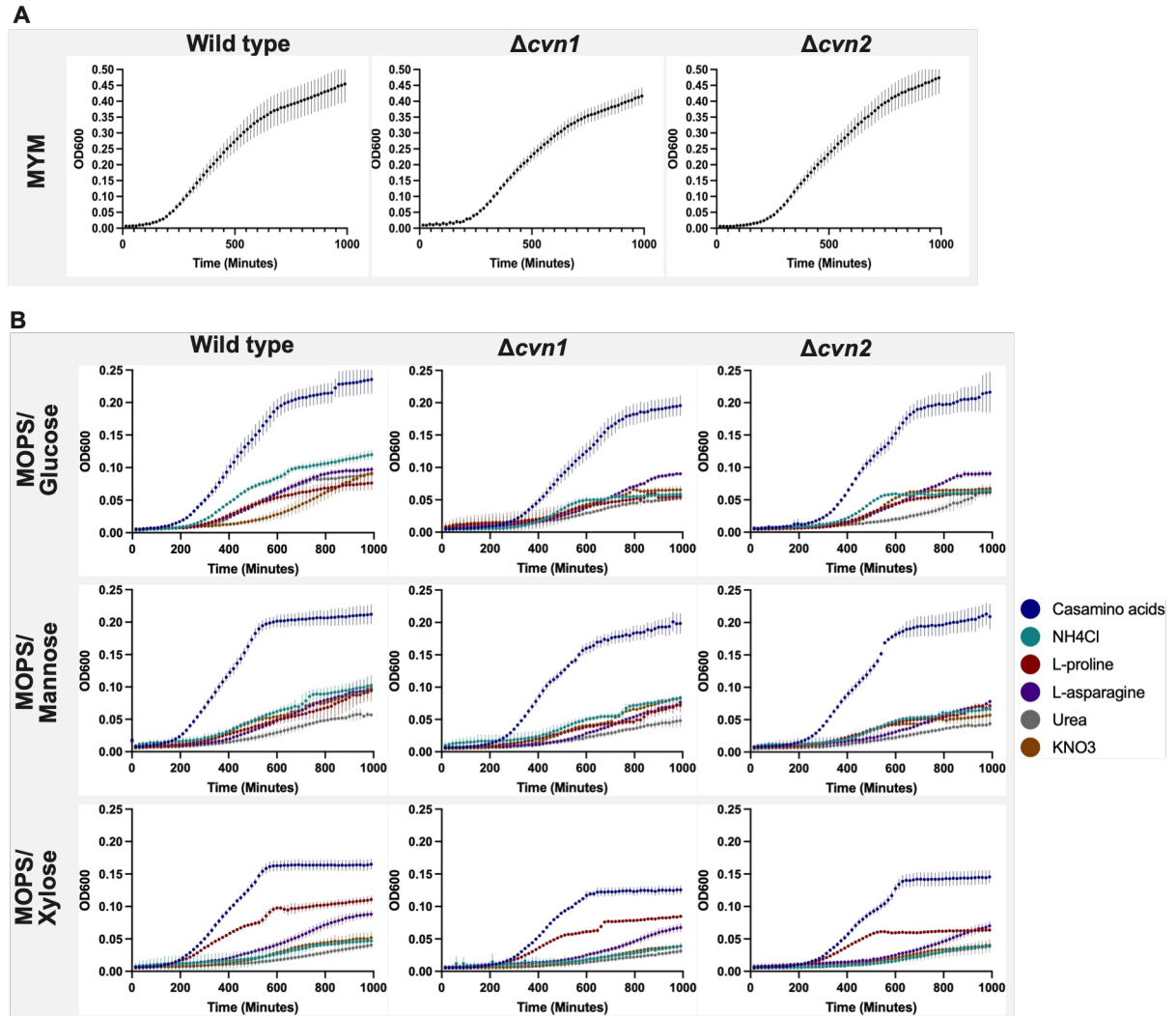


Figure S5. Morphological defect and lack of mycelial organization on poor media in *cvn1* and *cvn2* mutants. (A) Representative phase-contrast micrographs of vegetative hyphae of *S. venezuelae* wild type, *cvn1* mutant (LUV258), and *cvn2* mutant (LUV309) strains growing on cellophane membranes on MOPS minimal agar media containing all different combinations of the three carbon sources, glucose, xylose, and mannose, with the six nitrogen sources, ammonium chloride, potassium nitrate, urea, L-proline, L-asparagine, and casamino acids, for 13 hours. Scale bars, 20 μ m.



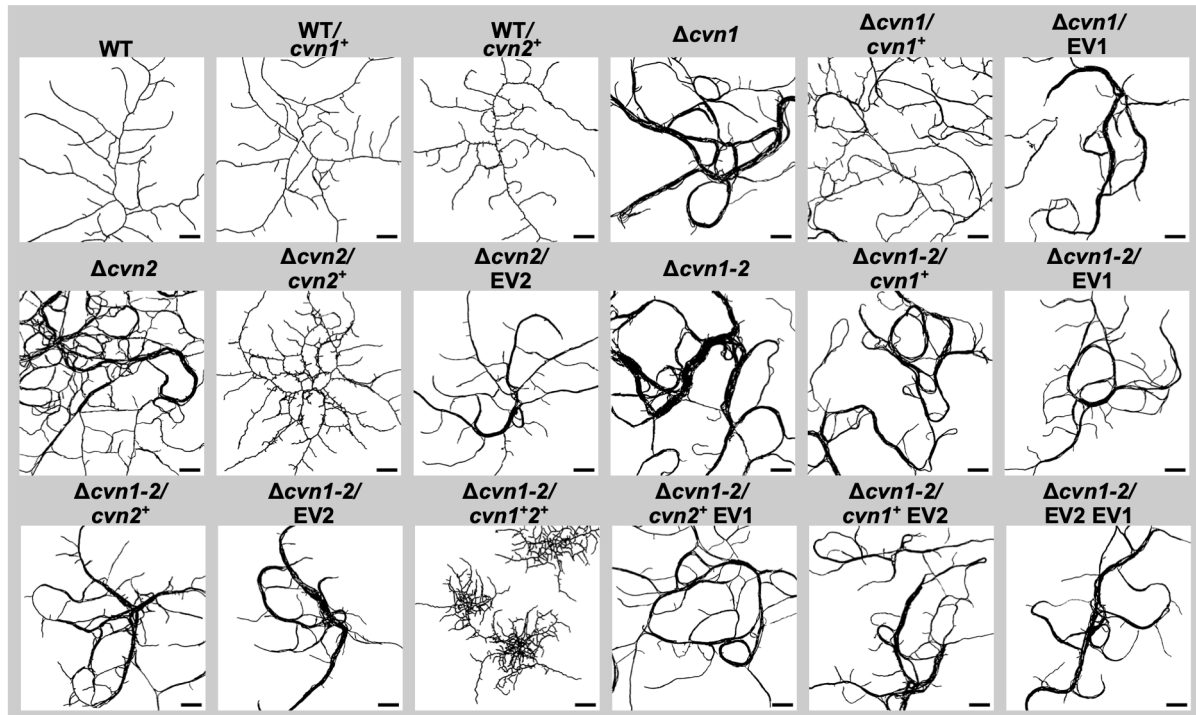


Figure S7. Control experiments of complementation tests of hyphal bundling phenotype in *cvn1* and *cvn2* deletion mutants. Representative segmented phase-contrast micrographs of vegetative hyphae of control strains constructed for complementation tests shown in Figure 4, and their parent wild-type and mutant strains. All strains were grown on cellophane on chitin agar for 13 hours. Scale bars 20 μ m. Strains shown are *S. venezuelae* wild type, LUV258 ($\Delta cvn1$), LUV309 ($\Delta cvn2$), LUV256 ($\Delta(cvn1-cvn2)$), and their corresponding complemented strains harboring plasmids pKF796 (*cvn1*⁺), pKF702 (*cvn2*⁺), pOSV804 (EV1: empty vector of plasmid pKF796 (*cvn1*⁺)), and pSS76 (EV2: empty vector of plasmid pKF702 (*cvn2*⁺)).

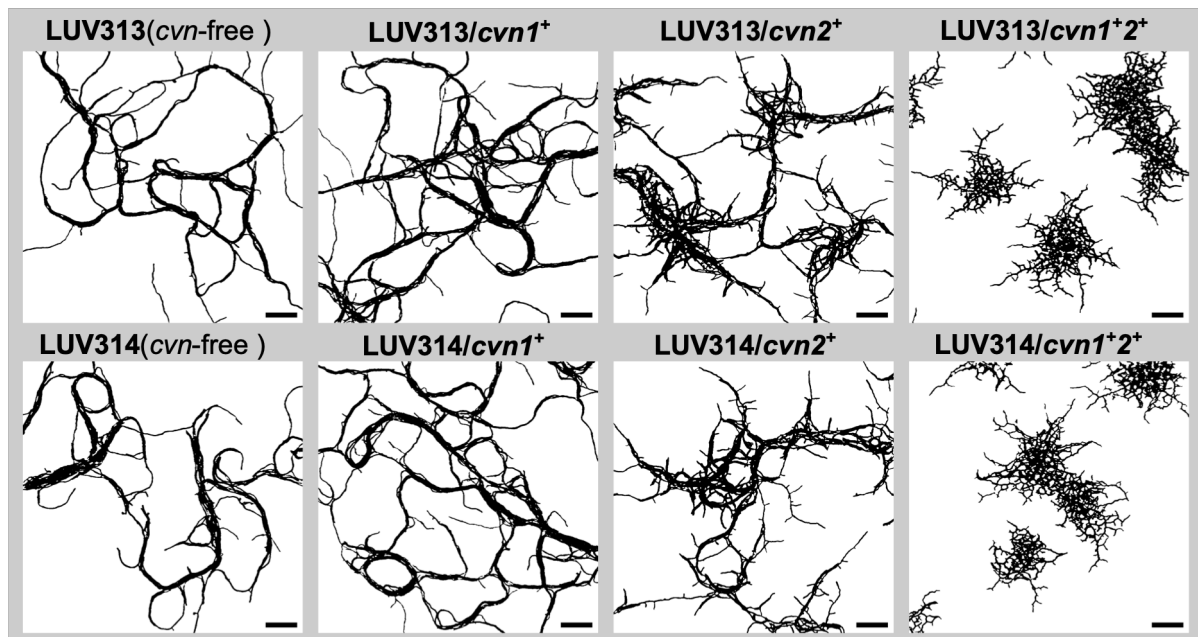


Figure S8. Both *cvn1* and *cvn2* are required to reverse hyphal bundling phenotype in the *cvn*-free mutant strains Representative segmented phase-contrast micrographs of vegetative hyphae of *cvn*-free mutant strains LUV313 and LUV314, and their corresponding complemented strains having either of or both plasmids pKF796 (*cvn1*⁺) and pKF702 (*cvn2*⁺). Note that pKF702 likely causes overexpression of *cvn2* from the *kasOp*^{*} promoter, appearing to affect the hyphal morphology. Strains were grown on cellophane membranes on chitin agar and phase-contrast imaging was done after 13 hours of incubation. Scale bars 20 μ m.

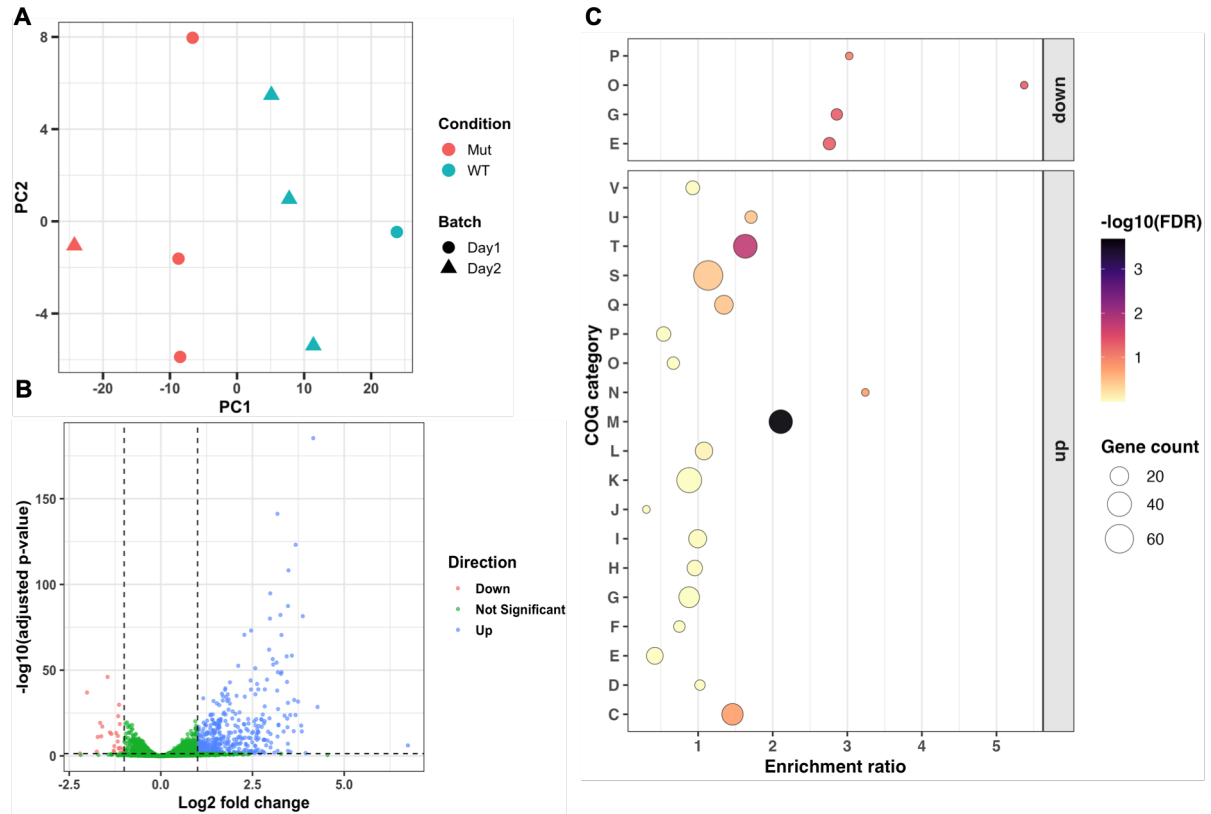


Figure S9. Transcriptome analysis of *cvn1-cvn2* mutant versus wild type. (A) PCA plot showing how samples cluster based on overall transcriptome similarity after normalization and batch adjustment. PCA was performed on batch-corrected VST-normalized counts, where batch effects were removed by modeling and subtracting batch-associated variation. Colors indicate conditions (mutant vs. WT) and shapes distinguish sequencing days (Day 1 vs. Day 2). (B) Volcano plot of differentially expressed genes filtered by adjusted p value of 0.05 and log2 fold change of 1. Up-regulated genes are shown in blue, down-regulated genes are shown in red, and non-significant genes are shown in green. (C) Bubble plot of COG enrichment analysis using the filtered differentially expressed gene set using adjusted p value of 0.05 and log2 fold change of 1, with COG categories on the y-axis, enrichment ratio on the x axis calculated as $(k/n)/(K/N)$ where k is the number of genes in the upregulated or downregulated gene set assigned to a given COG category, n is the total number of annotated genes in that gene set, K is the number of background genes assigned to the same COG category, and N is the total number of annotated background genes. Bubble size represents gene counts and colors indicate $-\log_{10}(\text{FDR})$. COG enrichment was assessed separately for up-regulated and down-regulated gene sets by comparing the frequency of genes assigned to each COG category against the annotated background using Fisher's exact test (one-sided overrepresentation test, alternative = "greater"). p values were adjusted using the Benjamini-Hochberg false discovery rate (FDR) method.

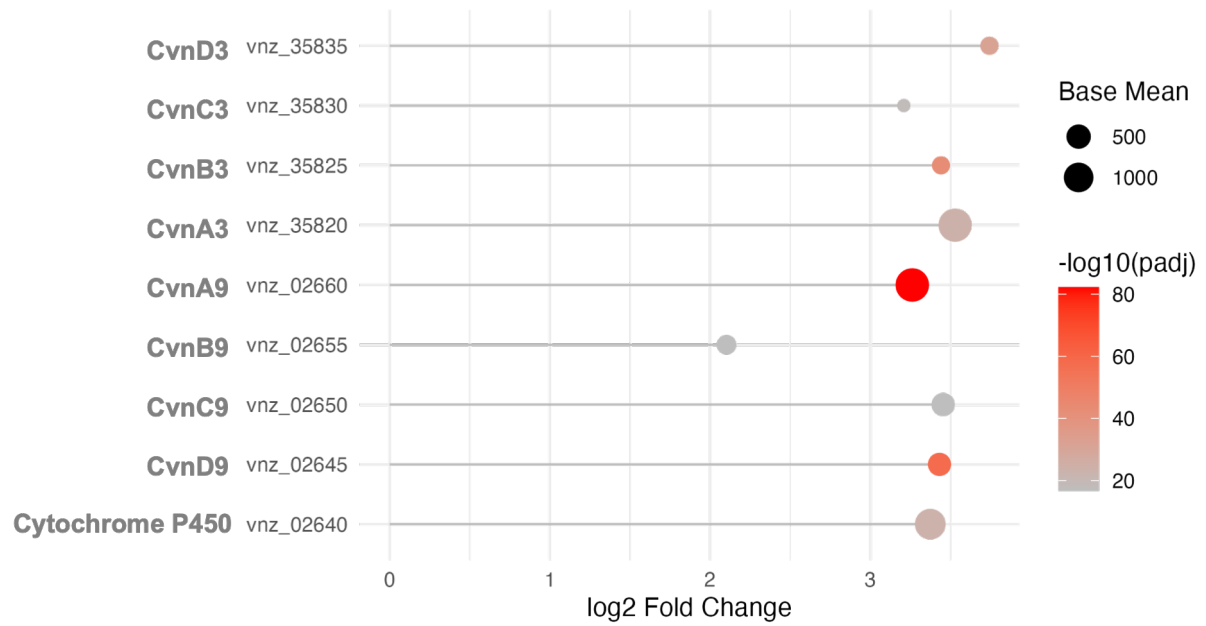


Figure S10. Conservons 3 and 9 are upregulated in *cvn1-cvn2* deletion mutant. Lollipop plot showing the upregulated conserved genes in *cvn1-cvn2* mutant compared to wild type. Log2 fold change is shown on the x-axis, dot color represents $-\log_{10}(\text{adjusted } p \text{ value})$, and dot size reflects “baseMean” to indicate overall expression level.

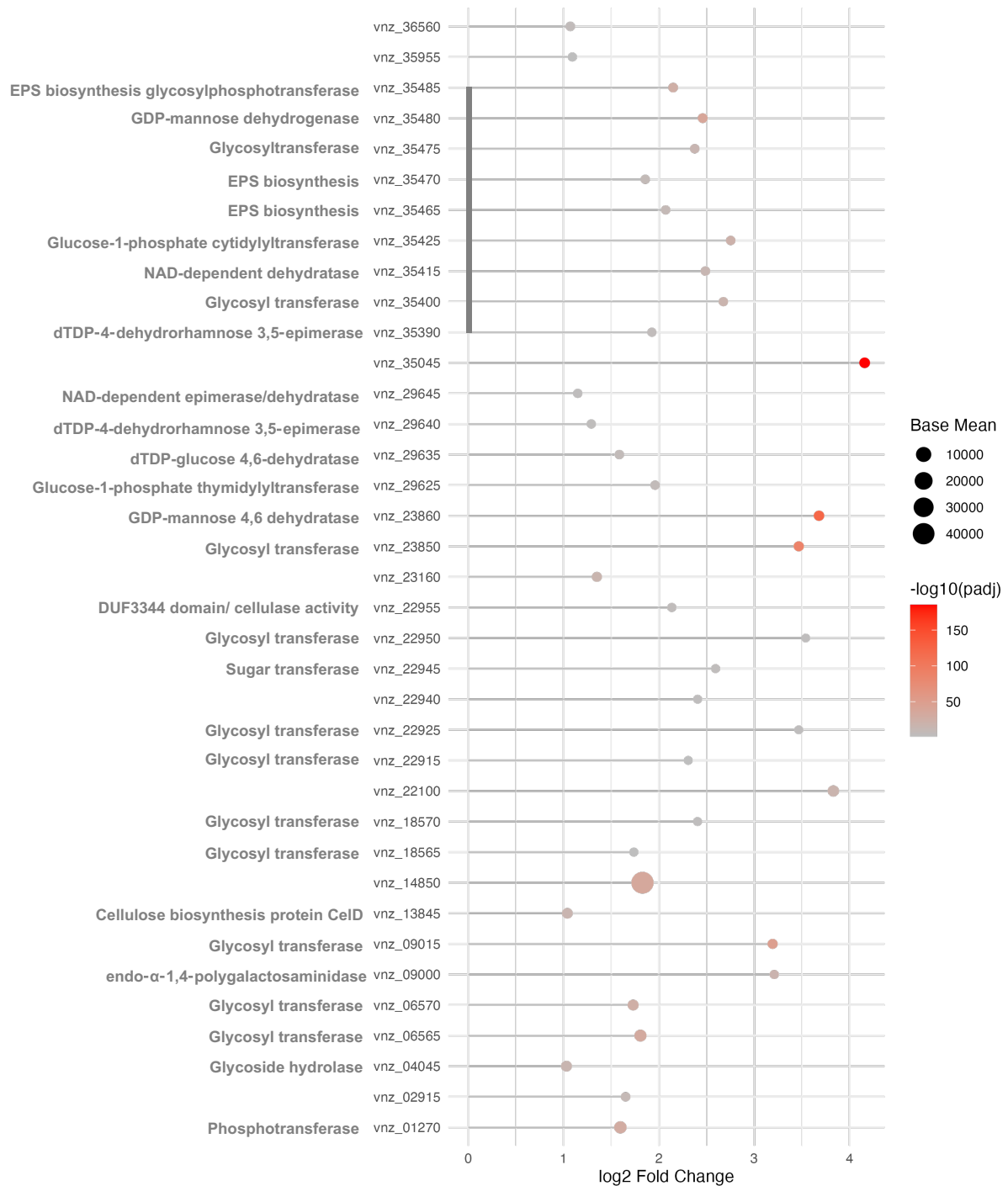


Figure S11. Upregulated genes under COG category M in the *cvn1-cvn2* mutant. Lollipop plot for all significantly upregulated genes under COG category M, filtered using adjusted p value of 0.05 and log2 fold change of 1. Log2 fold change is shown on the x-axis, dot color represents $-\log_{10}(\text{adjusted } p \text{ value})$, and dot size reflects “baseMean” to indicate overall expression level.

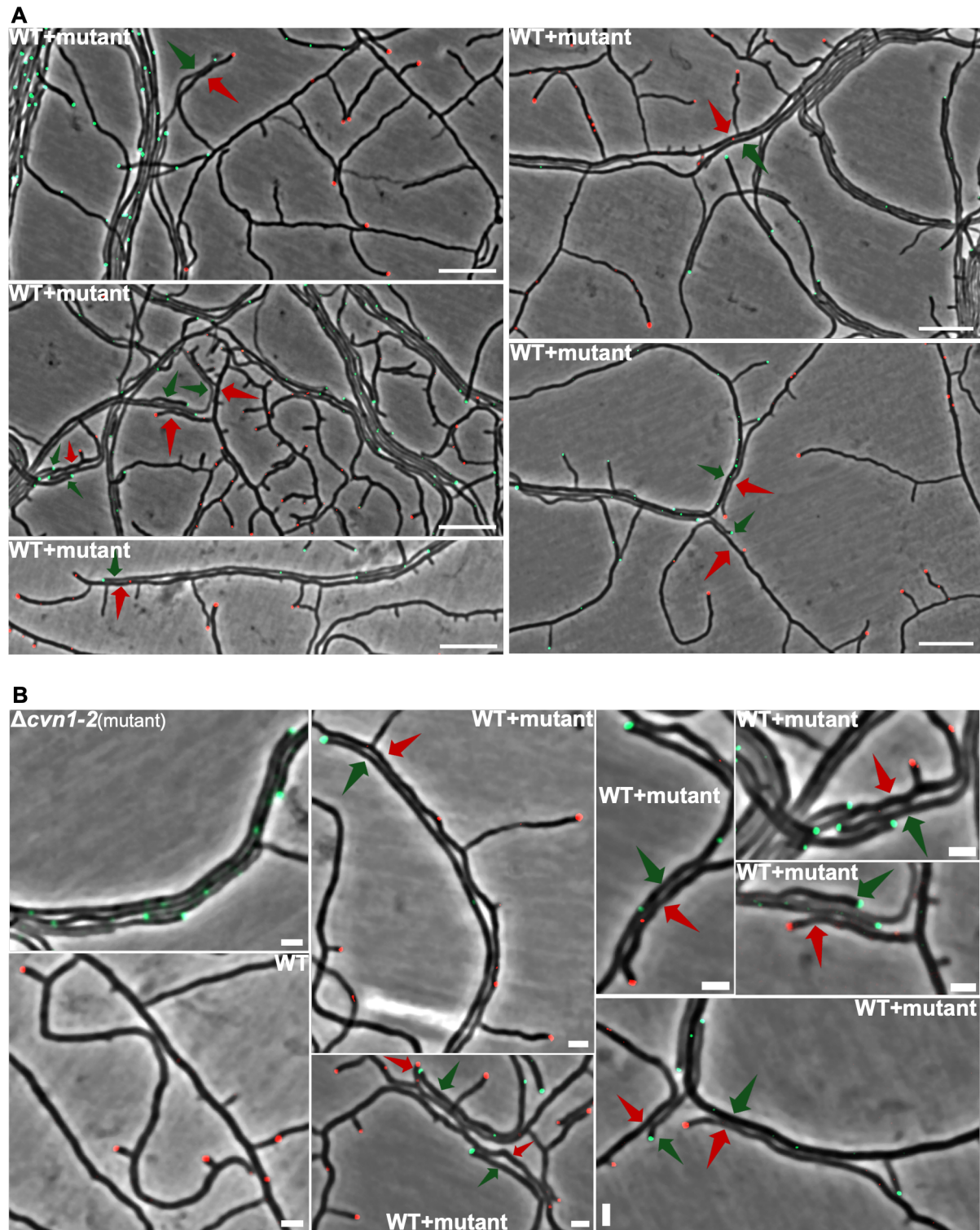


Figure S12. Hyphal bundling is driven by direct contact and not diffusible signals. (A) Representative merged phase-contrast and fluorescence micrographs of vegetative hyphae of co-cultured *S. venezuelae* wild type and LUV256 ($\Delta cvn1-cvn2$) mutant expressing DivIVA-mCherry (pSS204) and DivIVA-mNeonGreen (pKF731), respectively. **(B)** Close-up merged phase-contrast and fluorescence micrographs of co-cultured wild type and mutant, including regions corresponding to examples shown in panel (A) and Fig. 6, as well as additional representative regions not shown previously. In (A, B), red and green arrows point to wild-type and mutant hyphae, respectively, during

contact. Spores of strains were co-cultured 1.5 cm apart on the same cellophane on chitin agar and images were taken after 13 hours of growth. Scale bars, 10 μm in A and 2 μm in B

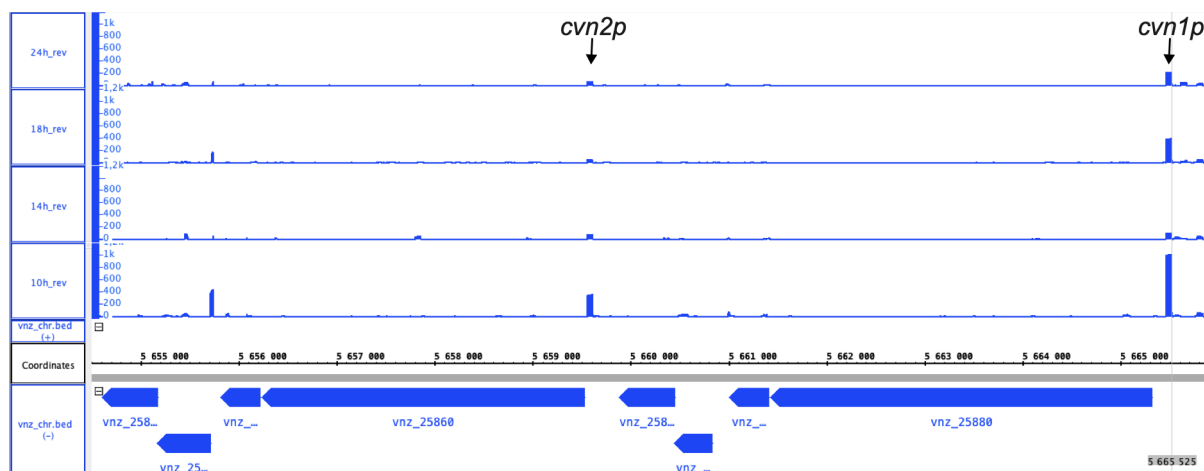


Figure S13. Transcription start site mapping in the *cvn1-cvn2* region of *S. venezuelae* genome. The figure is based on 5'-triphosphate end-capture RNA-seq data obtained by Matt Bush, John Innes Centre (ArrayExpress, accession number E-MTAB-10690) (https://streptomyces.org.uk/vnz_tss.html). Normalized coverage by TSS reads on the forward strand, subtracted by normalized coverage by control reads on the forward strand is displayed. The thin grey vertical line indicates the position of the estimated transcription start site for the *cvn1* operon that likely also contributes to expression of the downstream genes *cvn2* operon, even though there is a mapped start site also upstream of *cvn2*. As described, four time points along a liquid culture time course for *S. venezuelae* were analysed: 10 hr, 14 hr, 18 hr and 24 hr to represent the transition throughout development (vegetative, pre-sporulation, onset of sporulation and mid/late sporulation).

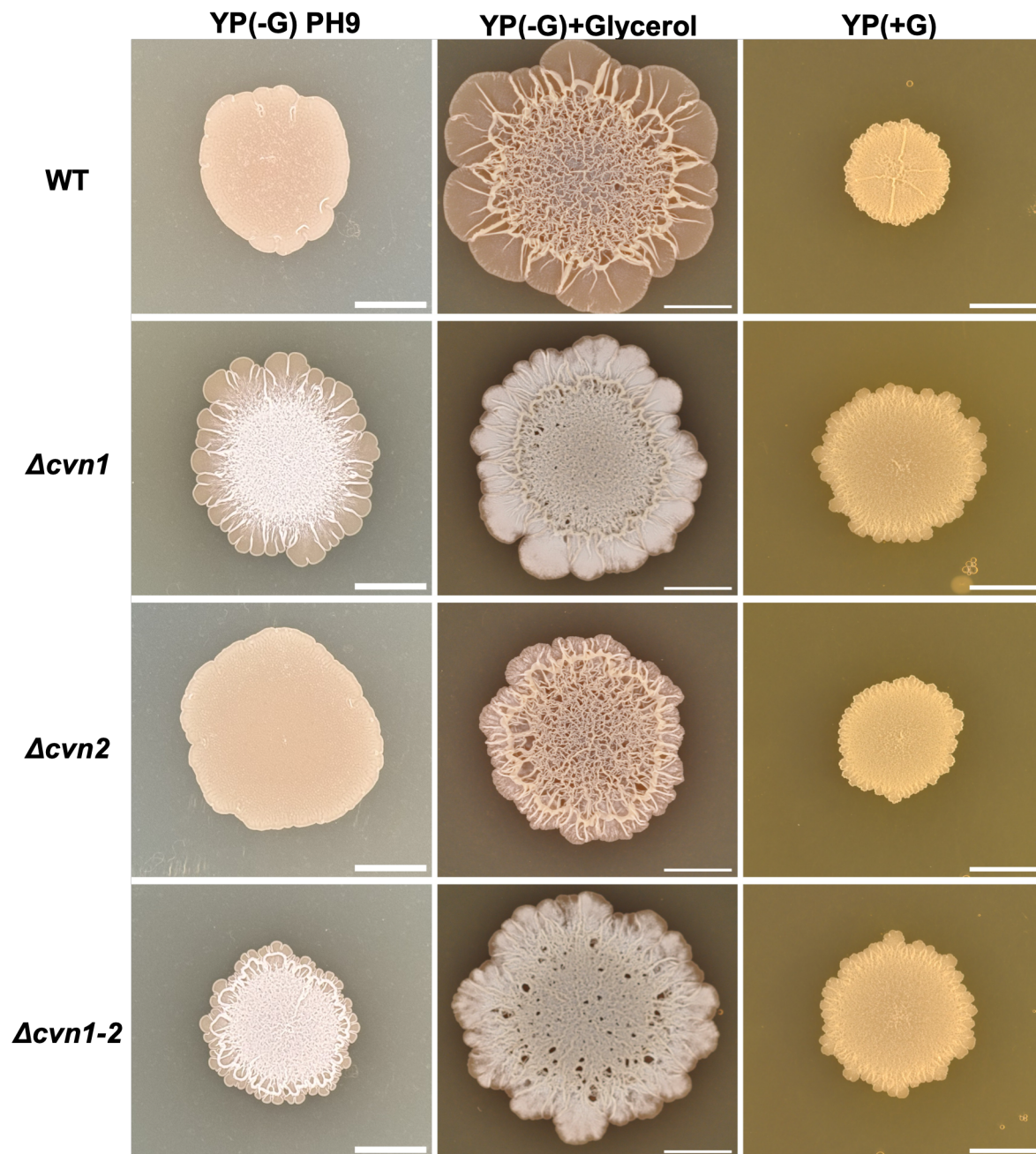


Figure S14. Exploratory behavior of mutants lacking either or both of *cvn1* and *cvn2* on different YP agar formulations. Representative images showing the exploratory behavior of the strains, from top to bottom, *S. venezuelae* wild type, LUV258 ($\Delta cvn1$), LUV309 ($\Delta cvn2$) and LUV256 ($\Delta(cvn1-cvn2)$) on different formulations of YP agar: from left to right, glucose-deficient (YP(-G)) adjusted to pH 9 after 4 days of growth, glucose-deficient (YP(-G)) supplemented with glycerol after 5 days of growth, and YP supplemented with glucose (YP(+G)) after 5 days of growth. 5 μ l of spore suspensions of similar concentrations of all strains were spotted on YP agar. Scale bars, 1 cm.

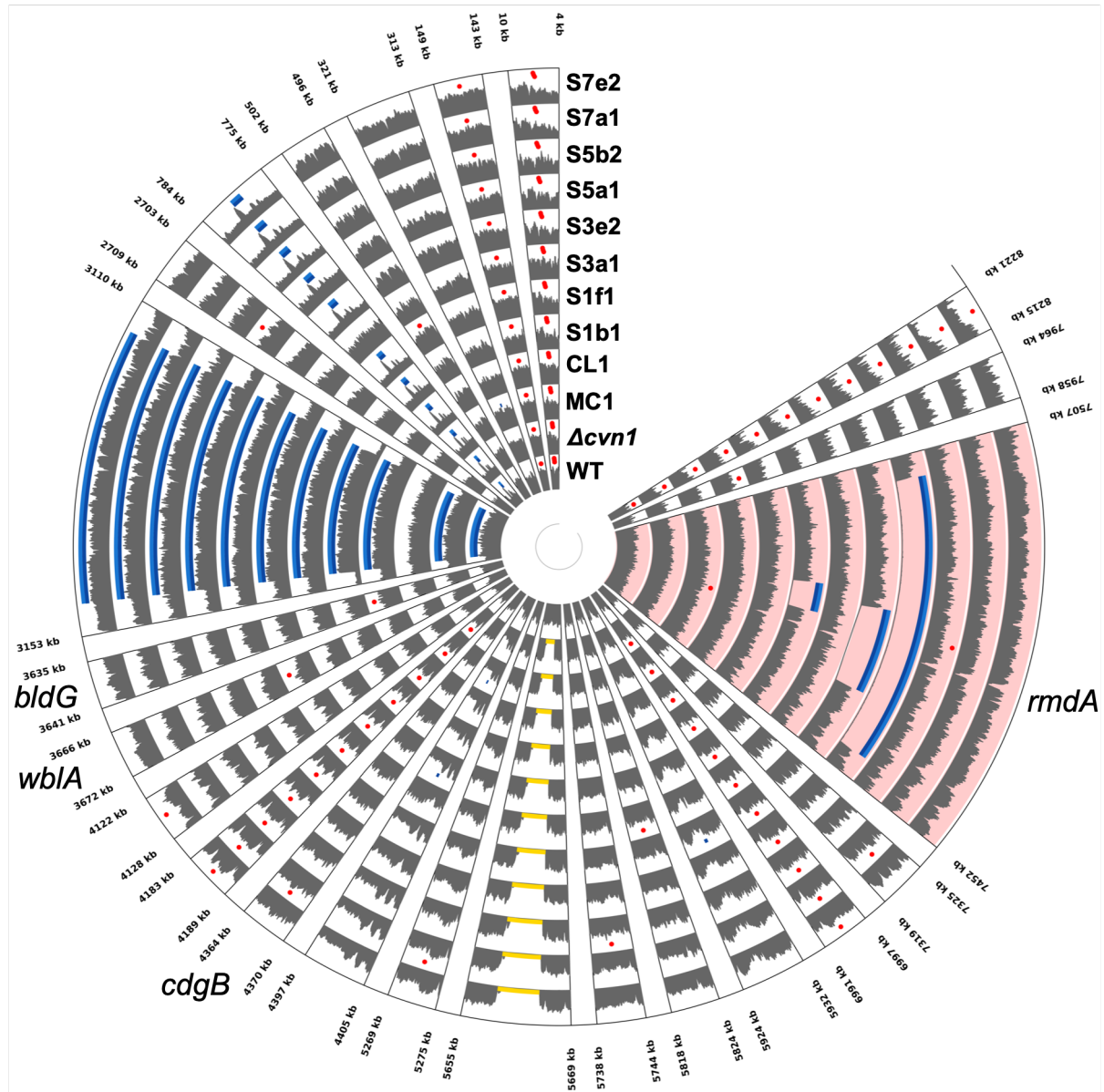


Figure S15. Summary of SNPs and SVs detected in genome sequences of $\Delta cvn1$ -suppressor strains. Circular genomic plot showing coverage of sequencing reads of the wild type, $\Delta cvn1$ (LUV258), and $\Delta cvn1$ -suppressor strains (CL1, MC1, S1b1, S1f1, S3a1, S3e2, S5a1, S5b2, S7a1, and S7e2) against *S. venezuelae* NRRL B-65442 wild-type reference genome. Only genomic regions containing either SNPs or SVs located in coding regions are plotted. SNPs are displayed in red. SVs are shown in light blue (DELLY-called variants), dark blue (LUMPY-called variants), or both when detected by both callers. *cvn1* deletion is highlighted in yellow, and the recurrently altered region that contains *rmdA* is shaded in red. Genome coordinates are indicated on the outer track. See also Supplementary Tables S12-S14 for full lists of SNPs and SVs. It is worth noting that despite filtering detected SVs, the plots and lists in supplementary tables S13 and S14 contain examples of variants reported in particular strains, although low-level evidence suggests they may also be present in additional strains.

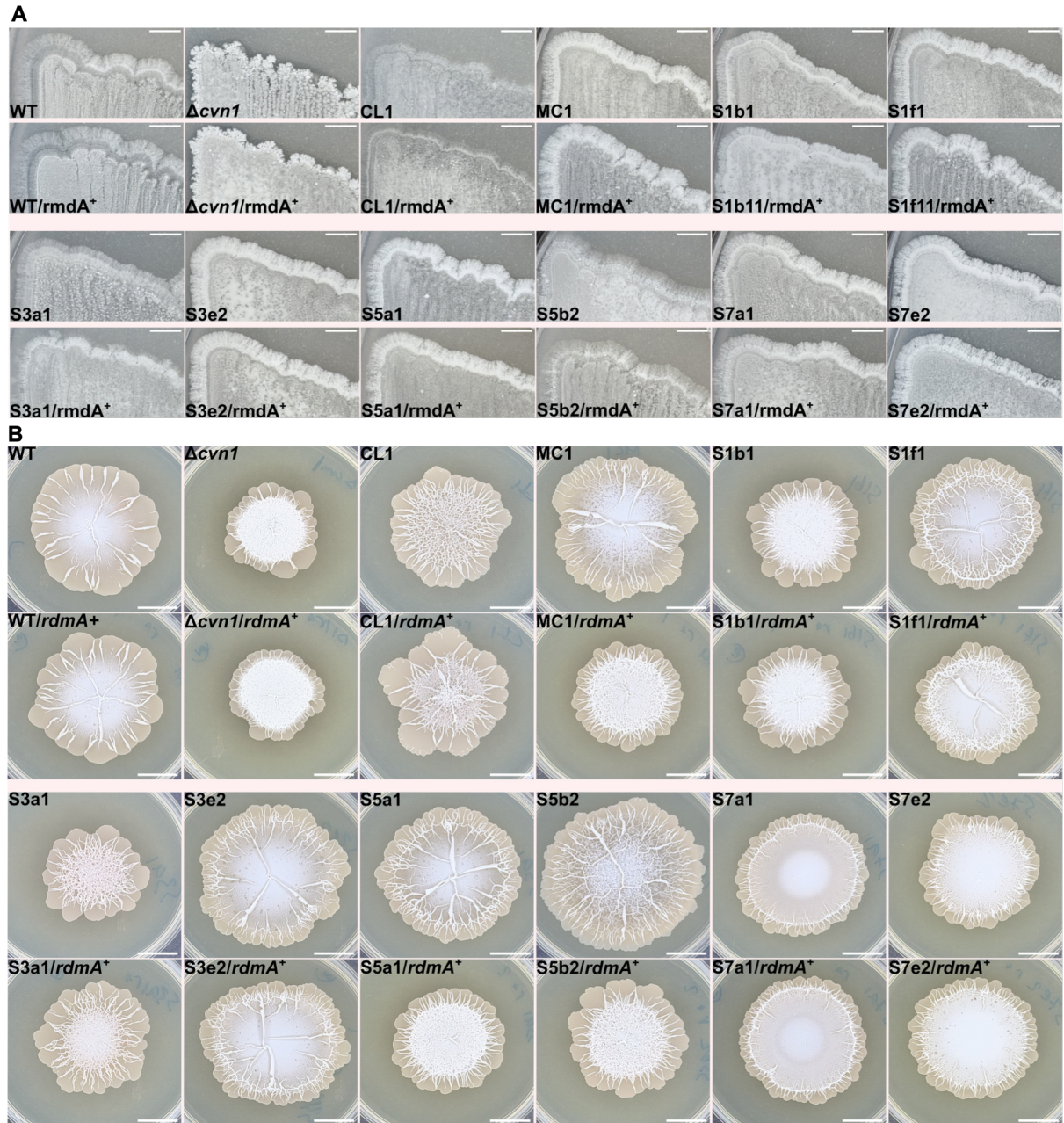


Figure S16. Complementation of $\Delta cvn1$ -suppressor strains with *rmdA* (A) Images showing the cropped edges of growth patches to visualize colony fringes of wild type, $\Delta cvn1$, and $\Delta cvn1$ -suppressor strains, and their corresponding strains with expression of *rmdA* *in trans* on chitin agar. Loopfuls of spore suspensions of similar concentrations were patch-streaked on chitin agar and imaging was done after 6 days of growth. Scale bars, 0.5 cm. (B) Images showing the exploratory growth phenotype of wild type, $\Delta cvn1$, and $\Delta cvn1$ -suppressors as well as their corresponding strains with expression of *rmdA* *in trans*. 5 μ l of spore suspensions of similar concentrations was spotted on YP(-G) and imaging was done after 6 days of growth. Scale bars, 1 cm. Strains shown are *S. venezuelae* wild type, LUV258 ($\Delta cvn1$), and the suppressor strains CL1, MC1, S1b1, S1f1, S3a1, S3e2, S5a1, S5b1, S5b2, S7a1, and S7e2, with and without the integrating plasmid pSVJH101 (*rmdA*⁺).

Supplementary Tables

Table S1. List of strains used in the study.

Strain	Genotype/Relevant characteristics	Source/reference
<i>Escherichia coli</i>		
DH5α	<i>supE44 ΔlacU169 (Φ80 lacZΔM15)</i> <i>hsdR17 recA1 endA1 gyrA96 thi-1</i> <i>relA1</i>	34
BT340	DH5α carrying FLP recombination plasmid pCP20 (<i>flp, bla, cat, rep101^{ts}</i>)	35
DY380	F ⁻ <i>mcrA</i> Δ(<i>mrr-hsdRMS-mcrBC</i>) <i>Φ80dlacZM15 ΔlacX74 deoR recA1</i> <i>endA1 araD139 Δ(ara, leu)7649 galU</i> <i>galK rspL nupG [λcl857 (cro-</i> <i>bioA)<>tet]</i>	36
ET12567/pUZ8002	<i>dam-13::Tn9 dcm-6 hsdM</i> , carrying helper plasmid pUZ8002	37
<i>Streptomyces venezuelae</i>		
NRRL B-65442	Wild type <i>S. venezuelae</i> strain	38
LUV256	Δ(<i>cvn1-cvn2</i>)::FRT	This work
LUV258	Δ <i>cvn1</i> ::FRT	This work
LUV259	Δ <i>cvn3</i> ::FRT	This work
LUV260	Δ <i>cvn7</i> ::FRT	This work
LUV263	Δ <i>cvn9</i> ::FRT	This work
LUV267	Δ(<i>cvn1-cvn2</i>)::FRT Δ <i>cvn3</i> ::FRT	This work
LUV270	Δ(<i>cvn1-cvn2</i>)::FRT Δ <i>cvn3</i> ::FRT Δ <i>cvn9</i> ::FRT	This work
LUV275	Δ(<i>cvn1-cvn2</i>)::FRT Δ <i>cvn3</i> ::FRT Δ <i>cvn9</i> ::FRT Δ <i>cvn5</i> ::FRT	This work
LUV276	Δ <i>cvn5</i> ::FRT	This work
LUV280	Δ(<i>cvn1-cvn2</i>)::FRT Δ <i>cvn3</i> ::FRT Δ <i>cvn9</i> ::FRT Δ <i>cvn5</i> ::FRT Δ <i>cvn7</i> ::FRT	This work
LUV309	Δ <i>cvn2</i> :: <i>apr^I</i>	This work
LUV310	Δ <i>cvn4</i> ::FRT	This work
LUV311	Δ <i>cvn6</i> ::FRT	This work
LUV312	Δ(<i>cvn1-cvn2</i>)::FRT Δ <i>cvn3</i> ::FRT Δ <i>cvn9</i> ::FRT Δ <i>cvn5</i> ::FRT Δ <i>cvn7</i> ::FRT Δ <i>cvn4</i> ::FRT	This work
LUV313	Δ(<i>cvn1-cvn2</i>)::FRT Δ <i>cvn3</i> ::FRT Δ <i>cvn9</i> ::FRT Δ <i>cvn5</i> ::FRT Δ <i>cvn7</i> ::FRT Δ <i>cvn4</i> ::FRT Δ <i>cvn6</i> ::FRT	This work
LUV314	Δ(<i>cvn1-cvn2</i>)::FRT Δ <i>cvn3</i> ::FRT Δ <i>cvn9</i> ::FRT Δ <i>cvn5</i> ::FRT Δ <i>cvn7</i> ::FRT Δ <i>cvn4</i> ::FRT Δ <i>cvn6</i> ::FRT	This work
CL1	Suppressor strain in <i>cvn1</i> mutant background	This work

MC1	Suppressor strain in <i>cvnI</i> mutant background	This work
S1b1	Suppressor strain in <i>cvnI</i> mutant background	This work
S1f1	Suppressor strain in <i>cvnI</i> mutant background	This work
S3a1	Suppressor strain in <i>cvnI</i> mutant background	This work
S3e2	Suppressor strain in <i>cvnI</i> mutant background	This work
S5a1	Suppressor strain in <i>cvnI</i> mutant background	This work
S5b2	Suppressor strain in <i>cvnI</i> mutant background	This work
S7a1	Suppressor strain in <i>cvnI</i> mutant background	This work
S7e2	Suppressor strain in <i>cvnI</i> mutant background	This work

¹*apr* denotes here denotes here the cassette containing both the apramycin resistance gene *aac(3)IV* and an *oriT* that derives from plasmid pIJ773.

Table S2. List of plasmids used in this study.

Plasmid	Description ¹	Source
pIJ773	Source of FRT- <i>oriT</i> - <i>aac(3)IV</i> -FRT cassette, Apra ^r	³⁹
pIJ10701	Source of <i>bla-oriT-hyg-bla</i> ; hygromycin resistance cassette with flanking <i>bla</i> sequencing to replace <i>bla</i> in cosmids	³⁹
pKF537	Synthesized mNeonGreen gene with codon choice optimized for <i>Streptomyces</i> ; Amp ^r	K. Flärdh
pKF702	<i>kasOp</i> *- <i>cvn2</i> for expression of <i>cvn2</i> <i>in trans</i> . Derivative of pSS76	This work
pKF712	Derivative of SuperCos 2F13, with $\Delta cvn5::apra$	This work
pKF714	Derivative of SuperCos Sv-6-G04, with $\Delta cvn7::apra$ T	This work
pKF715	Derivative of SuperCos Sv-4-G03, with $\Delta cvn9::apra$	This work
pKF720	Derivative of SuperCos 5-D08 with $\Delta (cvn1-cvn2)::apra$	This work
pKF721	Derivative of SuperCos 5-D08 with $\Delta cvn2::apra$	This work
pKF723	Derivative of SuperCos 5-D08 with $\Delta cvn1::apra$	This work
pKF724	Derivative of SuperCos 3L14 with $\Delta cvn3::apra$	This work
pKF725	Derivative of cosmid pKF712 $\Delta cvn5::FRT$	This work
pKF727	Derivative of cosmid pKF715 $\Delta cvn9::FRT$	This work
pKF728	Derivative of cosmid pKF714 with $\Delta cvn7::FRT$	This work
pKF729	Derivative of cosmid pKF723 with $\Delta cvn1::FRT$	This work
pKF730	Derivative of cosmid pKF724 with $\Delta cvn3::FRT$	This work
pKF731	<i>divIVA-mNeonGreen</i> ; derived from pSS204 with <i>mNeongreen</i> from pKF537	This work
pKF735	Derivative of cosmid pKF729 ($\Delta cvn1::FRT$) with <i>hyg-oriT</i>	This work
pKF736	Derivative of cosmid pKF730 ($\Delta cvn3::FRT$) with <i>hyg-oriT</i>	This work

pKF737	Derivative of cosmid pKF725 ($\Delta cvn5::FRT$) with <i>hyg-oriT</i>	This work
pKF738	Derivative of cosmid pKF728 ($\Delta cvn7::FRT$) with <i>hyg-oriT</i>	This work
pKF739	Derivative of cosmid pKF727 ($\Delta cvn9::FRT$) with <i>hyg-oriT</i>	This work
pKF740	Derivative of cosmid pKF720 with $\Delta(cvn1-cvn2)::FRT$	This work
pKF796	P _{<i>cvn1-cvn1</i>} for expression of <i>cvn1 in trans</i> . Derivative of pOSV804	This work
pKF823	Derivative of cosmid pKF740 $\Delta(cvn1-cvn2)::FRT$ with <i>hyg-oriT</i>	This work
pKF873	Derivative of SuperCos Pl-2-J6 with $\Delta cvn6::apra$	This work
pKF874	Derivative of SuperCos Pl-2-M21 with $\Delta cvn4::apra$	This work
pKF875	Derivative of pKF873 with $\Delta cvn6::FRT$	This work
pKF877	Derivative of pKF874 with $\Delta cvn4::FRT$	This work
pKF879	Derivative of pKF875 ($\Delta cvn6::FRT$) with <i>hyg-oriT</i>	This work
pKF881	Derivative of pKF877 ($\Delta cvn4::FRT$) with <i>hyg-oriT</i>	This work
pOSV804	Vector that integrates via <i>attP</i> VWB, Apra ^r	40
pSS76	Vector with <i>kasOp*</i> that integrates via <i>attP</i> ϕ BT1, Hyg ^r	Susan Schlimpert, John Innes Centre, Norwich, UK
pSS204	pIJ10770 containing <i>divIVA-mCherry</i> driven by the <i>divIVA</i> promoter; integrates via <i>attP</i> ϕ BT1; Hyg ^r	Susan Schlimpert, John Innes Centre, Norwich, UK
pSVJH101	<i>rmdA-FLAG</i> in derivative of pMS81; integrates via <i>attP</i> ϕ BT1; Hyg ^r	Natalia Tschowri, Leibniz Universität, Hannover
2F13	Cosmid with chromosomal region of <i>S. venezuelae cvn5</i> . Amp ^r Kan ^r	StrepDB database ²
3L14	Cosmid with chromosomal region of <i>S. venezuelae cvn3</i> . Amp ^r Kan ^r	StrepDB database ²
5-D08	Cosmid with chromosomal region of <i>S. venezuelae cvn1</i> and <i>cvn2</i> . Amp ^r Kan ^r	StrepDB database ²

Sv-4-G03	Cosmid with chromosomal region of <i>S. venezuelae</i> cvn9. Amp ^r Kan ^r	StrepDB database ²
Sv-6-G04	Cosmid with chromosomal region of <i>S. venezuelae</i> cvn7. Amp ^r Kan ^r	StrepDB database ²
Pl-2-J6	Cosmid with chromosomal region of <i>S. venezuelae</i> cvn6. Amp ^r Kan ^r	StrepDB database ²
Pl-2-M21	Cosmid with chromosomal region of <i>S. venezuelae</i> cvn4. Amp ^r Kan ^r	StrepDB database ²

¹ Amp^r, ampicillin resistance; Apra^r, apramycin resistance; Hyg^r, Hygromycin resistance

² <https://strepdb.streptomyces.org.uk>

Table S3. List of primers used in the study.

Oligos	Sequence (5'→3') ¹	Description
KF849	CCATTCTTCAGGATGGCAAG	Primer inside <i>apra</i> cassette. Used for verification of gene replacement by <i>aac(3)IV-oriT</i>
KF1128	TGCAACAGTGCCGTTGATC	Primer for verification of <i>hyg-oriT</i> integration into cosmid DNA
KF1246	CCTCTGACCCCTGACCCCGTC	Sequencing primer for verification of cloning of <i>cvn2</i> in pSS76
KF1272	GAGTTGGTAGCTCTTGATCCGGC	Primer for verification of <i>hyg-oriT</i> integration into cosmid DNA
KF1396	tcctctccgttcacccggaaggacgcctggccagaggttgcg acgctaATTCCGGGGATCCGTCGACC	Forward primer for λRed recombineering of <i>cvn1-cvn2</i> . Used in combination with KF1399
KF1397	cgaacaccggcacaacgtaggtgccgacgccgagaggttg ttggtcgaATTCCGGGGATCCGTCGACC	Forward primer for λRed recombineering of <i>cvn2</i> . Used in combination with KF1399
KF1399	cgctccccgggggaacgcggccttcggtacgtggacgccg ggggacagTGTAGGCTGGAGCTGCTTC	Reverse primer for λRed recombineering of <i>cvn1-cvn2</i> , when used with KF1396, or of <i>cvn2</i> when used with KF1397
KF1404	GGATCCTGGTGGCGGTTTC	Primer used in combination with KF1461 to verify $\Delta(cvn1-cvn2)::FRT$
KF1404	GGATCCTGGTGGCGGTTTC	Primer used in combination with KF1406 for verification of <i>cvn1-cvn2</i> or <i>cvn2</i> replacement by <i>aac(3)IV-oriT</i>
KF1406	ACCATTCCTTCAGGATGGCAAG	Primer inside <i>apra</i> cassette. Used for verification of gene replacement by <i>aac(3)IV-oriT</i>
KF1407	GACTCGTCGCCATCCACC	Primer used in combination with KF1404 for verification of the loss of wild type allele in <i>cvn1-cvn2</i> or <i>cvn2</i> mutants
KF1521	ctgatgcgatcatatgGTGAGGCGAAGCAAGGA AAG	Forward primer for amplification of <i>cvn2</i> for cloning into pSS76 for complementation tests
KF1522	ctgatgcgatcctaggTCACGCCGAAGGCCGCG CTC	Reverse primer for amplification of <i>cvn2</i> for cloning into pSS76 for complementation tests
KF1545	TCCTGAAGAATGGTGCAGTG	Primer inside <i>apra</i> cassette. Used for verification of gene replacement by <i>aac(3)IV-oriT</i>
KF1629	cgtcacctctgctttgtggcgacgagtcagaatgcgagca ccggtgTGTAGGCTGGAGCTGCTTC	Forward primer for λRed recombineering of <i>cvn5</i>
KF1630	ttccgcatgggtggtgctcctggcgtgacggggcgagctgag cgatctaATTCCGGGGATCCGTCGACC	Reverse primer for λRed recombineering of <i>cvn5</i>

KF1631	ccccgcagtacgccgattcctgggcacgtacgtaccaaggac actcatgTGTAGGCTGGAGCTGCTTC	Forward primer for λ Red recombineering of <i>cvn7</i>
KF1632	tccttgctgacgggggtcagcaggaggtgaccggtcgggtcg tacgtcaATTCCGGGGATCCGTCGACC	Reverse primer for λ Red recombineering of <i>cvn7</i>
KF1633	gaacgtgaagacccccgtggtcacggggccgagctgggg gaggccatgTGTAGGCTGGAGCTGCTTC	Forward primer for λ Red recombineering of <i>cvn4</i>
KF1634	tcggcgggtggtggcggtccgctgagacgggctcggcgggcg ggatatcaATTCCGGGGATCCGTCGACC	Reverse primer for λ Red recombineering of <i>cvn4</i>
KF1635	ccgcgaacacatgtgatcacgacgtgattaatatgcacgcgtg gcgtgTGTAGGCTGGAGCTGCTTC	Forward primer for λ RED recombineering of <i>cvn6</i>
KF1636	gacgggatggggcgagggtggaggcaggcgggcgggcc ttgttcgtcATTCCGGGGATCCGTCGACC	Reverse primer for λ Red recombineering of <i>cvn6</i>
KF1637	caggatccaccgcaccagagccagtcgcagccggaacgag gaggccatgTGTAGGCTGGAGCTGCTTC	Forward primer for λ Red recombineering of <i>cvn3</i>
KF1638	ggctcgtcgcccaggccggcggtggccgactgaggcccgcgcg gcggctcaATTCCGGGGATCCGTCGACC	Reverse primer for λ Red recombineering of <i>cvn3</i>
KF1639	cggcgacttcagtcgcctcattcctgactgttaggaccatt gatgTGTAGGCTGGAGCTGCTTC	Forward primer for λ Red recombineering of <i>cvn9</i>
KF1640	gggaccgcccctgccttctcggcggtgtgttcggcgggtggg agggtcaATTCCGGGGATCCGTCGACC	Reverse primer for λ Red recombineering of <i>cvn9</i>
KF1641	tctcgttcacccggaaggacgcctggtccagaggttgcgacg ctagtTGTAGGCTGGAGCTGCTTC	Forward primer for λ Red recombineering of <i>cvn1</i>
KF1642	agaccgccgctggcgacggcatatgacaaccggcgagg tcgggtaATTCCGGGGATCCGTCGACC	Reverse primer for λ Red recombineering of <i>cvn1</i>
KF1664	ACGGGTCAGCTTGGTGAG	Sequencing primer for verification of cloning of <i>cvn1</i> in pOSV804
KF1669	CACCGGACCTACAAATGACC	Primer used in combination with KF1670 to verify the loss of wild type allele in <i>cvn5</i> mutants
KF1670	GATCTGGCGCTTGTGCTTG	Primer used in combination with KF849 for verification of <i>cvn5</i> replacement by <i>aac(3)IV-oriT</i>
KF1671	TCAGGATCTCCCACTCGAAC	Used in combination with KF1932 to verify $\Delta cvn6::FRT$
KF1673	CGCTCTACGGGTAAGTTTGC	Primer used in combination with KF849 for verification of <i>cvn7</i> replacement by <i>aac(3)IV-oriT</i>
KF1674	GACCGTGAGCAGTTCGTTC	Primer used in combination with KF1673 to verify the loss of wild type allele in <i>cvn7</i> mutants
KF1675	GGAGGTGATCGGTTTCATTGC	Primer used in combination with KF849 for verification of <i>cvn9</i> replacement by <i>aac(3)IV-oriT</i>
KF1676	TTCACGGAGCGCAGGATC	Primer used in combination with KF1675 to verify the loss of wild type allele in <i>cvn9</i> mutants

KF1711	GACGACCGGCACCATGAG	Primer used in combination with KF1710 to verify the loss of wild type allele in <i>cvn3</i> mutants
KF1719	tgatatttaaCTCGAGATGGTCTCCAAGGGC GAAG	Primer used to amplify <i>mNeongreen</i> from pKF537, with XhoI site
KF1720	tagagtcgtaGGTACCTCTCACTTGTACAGC TCGTCC	Primer used to amplify <i>mNeongreen</i> from pKF537, with KpnI site
KF1721	GTTCAACACGGGCAAGGAC	Primer used in combination with KF1710 to verify $\Delta cvn3::FRT$
KF1722	TGAATTGCAGCACGTGAAGG	Primer used in combination with KF1461 to verify $\Delta cvn1::FRT$
KF1723	AGAAGTTCAGGCCGAGGTC	Primer used in combination with KF1669 to verify $\Delta cvn5::FRT$
KF1724	GGATGCCGATCTCGTCCAC	Primer used in combination with KF1673 to verify <i>cvn7::FRT</i>
KF1725	CAGAACGTGTGATCAGCTG	Primer used in combination with KF1675 to verify $\Delta cvn9::FRT$
KF1737	TGGAAGTGGAGTGGTACGAG	Used in combination with KF1930 to verify $\Delta cvn4::FRT$
KF1738	GCCGCCGTTTGTGAATTC	Used in combination with KF1737 to verify $\Delta cvn4::FRT$
KF1864	tagagtcgatcttagaGCGATACATACCGGGCAT TAC	Forward primer for amplification of <i>cvn1</i> with native promoter for cloning into pOSV804
KF1865	ctgatgcgatcttaagCCCCAATTCTCTGTCGAA AC	Forward primer for amplification of <i>cvn1</i> with native promoter for cloning into pOSV804
KF1930	CCAAGTCCCTGAATAACAGG	Used in combination with KF849 of <i>cvn4</i> replacement by <i>aac(3)IV-oriT</i>
KF1931	CGAGGTAGATGAGGGACTGG	Used in combination with KF1930 to verify the loss of wild type allele in <i>cvn4</i> mutants
KF1932	TGGTGGAGCTGCTCACCACG	Used in combination with KF849 of <i>cvn6</i> replacement by <i>aac(3)IV-oriT</i>
KF1933	GCTGAAGTACACGAGGTTTCG	Used in combination with KF1932 to verify the loss of wild type allele in <i>cvn6</i> mutants
KF2009	AACCCTGACGTACATGGCTC	Used in combination with KF2010 to verify $\Delta cvn6::FRT$. Also used in combination with KF1933 to verify the loss of wild type allele in <i>cvn6</i> mutants
KF2010	ATCTCCGCGATCTCCACGGTG	Used in combination with KF2009 to verify $\Delta cvn6::FRT$
KF2167	GAGCGCCGCTGGAGAAAACC	For analysis of chromosomal deletion
KF2168	TGGGCCTACGGTGACCTTCTG	For analysis of chromosomal deletion
KF2172	GCACCCACGTACGCGATCTG	For analysis of chromosomal deletion
KF2175	TGGGGTGCCGATGTGTGTCC	For analysis of chromosomal deletion

N677	CCCTGATAAATGCTTCAATAATATTGAA AAAGGAAGAGTA	Forward primer for amplification of <i>bla-oriT-hyg-bla</i> from pIJ10701
N678	AATCAATCTAAAGTATATATGAGTAAAC TTGGTCTGACAG	Reverse primer for amplification of <i>bla-oriT-hyg-bla</i> from pIJ10701

¹The underlined regions denote different restriction sites. The lowercase denotes non-binding sequences.

Supplementary Files

Supplementary Tables S4-S14 are provided as Supplementary Files.

References

- 1 Snippy: Fast bacterial variant calling from NGS reads (2015).
- 2 Okonechnikov, K., Conesa, A. & García-Alcalde, F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292-294, doi:10.1093/bioinformatics/btv566 (2016).
- 3 Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* **15**, R84, doi:10.1186/gb-2014-15-6-r84 (2014).
- 4 Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333-i339, doi:10.1093/bioinformatics/bts378 (2012).
- 5 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv e-prints*, arXiv:1303.3997, doi:10.48550/arXiv.1303.3997 (2013).
- 6 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008, doi:10.1093/gigascience/giab008 (2021).
- 7 Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503-2505, doi:10.1093/bioinformatics/btu314 (2014).
- 8 Jeffares, D. C. *et al.* Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat. Commun.* **8**, 14061, doi:10.1038/ncomms14061 (2017).
- 9 Robinson, J. T. *et al.* Integrative genomics viewer. *Nature Biotechnol.* **29**, 24-26, doi:10.1038/nbt.1754 (2011).
- 10 Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**, 178-192, doi:10.1093/bib/bbs017 (2013).
- 11 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
- 12 pandas-dev/pandas: Pandas (2026).
- 13 Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357-362, doi:10.1038/s41586-020-2649-2 (2020).
- 14 Python v. 3.14.2 (2016).
- 15 A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisulfite-Seq) libraries. (2021).
- 16 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
- 17 BBMap: A Fast, Accurate, Splice-Aware Aligner. (2014).
- 18 FastQC: A Quality Control Tool for High Throughput Sequence Data (2010).
- 19 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).
- 20 Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048, doi:10.1093/bioinformatics/btw354 (2016).
- 21 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
- 22 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47, doi:10.1093/nar/gkv007 (2015).
- 23 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2021).
- 24 ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag New York, 2016).
- 25 data.table: Extension of `data.frame` (2026).
- 26 Wickham, H. Reshaping Data with the reshape Package. *Journal of Statistical Software* **21**, 1 - 20, doi:10.18637/jss.v021.i12 (2007).
- 27 dplyr: A Grammar of Data Manipulation (2026).
- 28 Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS One* **11**, e0163962, doi:10.1371/journal.pone.0163962 (2016).
- 29 Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* **38**, 5825-5829, doi:10.1093/molbev/msab293 (2021).

- 30 Huerta-Cepas, J. *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309-D314, doi:10.1093/nar/gky1085 (2019).
- 31 Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261-272, doi:10.1038/s41592-019-0686-2 (2020).
- 32 statsmodels: Econometric and statistical modeling with python (2010).
- 33 Wickham, H. *et al.* Welcome to the Tidyverse. *Journal of Open Source Software* **4**, 1686, doi:10.21105/joss.01686 (2019).
- 34 Hanahan, D. Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**, 557-580, doi:10.1016/s0022-2836(83)80284-8 (1983).
- 35 Cherepanov, P. P. & Wackernagel, W. Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of FLP-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158**, 9-14, doi:10.1016/0378-1119(95)00193-a (1995).
- 36 Lee, E. C. *et al.* A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73**, 56-65, doi:10.1006/geno.2000.6451 (2001).
- 37 Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. *Practical Streptomyces Genetics*. (The John Innes Foundation, 2000).
- 38 Gomez-Escribano, J. P. *et al.* *Streptomyces venezuelae* NRRL B-65442: Genome sequence of a model strain used to study morphological differentiation in filamentous actinobacteria. *J Ind Microbiol Biotechnol* **48**, doi:10.1093/jimb/kuab035 (2021).
- 39 Gust, B., Challis, G. L., Fowler, K., Kieser, T. & Chater, K. F. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci U S A* **100**, 1541-1546, doi:10.1073/pnas.0337542100 (2003).
- 40 Aubry, C., Pernodet, J. L. & Lautru, S. Modular and integrative vectors for synthetic biology applications in *Streptomyces* spp. *Appl. Environ. Microbiol.* **85**, e00485-00419, doi:<https://doi.org/10.1128/AEM.00485-19> (2019).