

Supplementary materials

Table S1. RAST subsystem-based functional annotation of the *Streptomyces albidoflavus* MUM256 genome.

Metric	Value
Total features	6,335
Coding genes (CDS)	6,162
RNA features	86
Repeat features	87
Features assigned to ≥ 1 subsystem	1,272
CDS assigned to ≥ 1 subsystem	1,224
Distinct individual subsystem terms	312
CDS containing hypothetical-protein designation	1,939

Table S2. The 165 publicly available genomes annotated as *S. albidoflavus* were retrieved from the NCBI assembly database.

Assembly Accession	Organism Infraspecific Names Strain	Assembly Level	Assembly Stats Number of Contigs	PGAP2 incomplete gene percentage (%)*
GCF_045039635.1	NBC_01110	Complete Genome	2	1.248768
GCF_000156475.1	J1074	Scaffold	501	12.63158
GCF_000359525.1	J1074	Complete Genome	1	1.132464
GCF_000717375.1	NRRL B-2307	Contig	135	3.511053
GCF_000719435.1	NRRL F-5618	Contig	165	3.669571
GCF_000719495.1	NRRL F-5621	Contig	240	6.155633
GCF_000719955.1	NRRL B-1271	Contig	115	2.702703
GCF_001509565.1	NRRL WC-3066	Contig	287	4.880893
GCF_001577385.1	SM254	Complete Genome	1	1.586523
GCF_001865795.1	OsiLf-2	Scaffold	54	1.722391
GCF_002289305.1	145	Scaffold	1199	27.59194
GCF_002289335.1	132	Contig	1053	25.17308
GCF_002317005.1	38	Contig	1114	23.80864
GCF_002317025.1	138	Contig	1028	23.18089
GCF_002317035.1	143	Contig	957	21.08148
GCF_004078575.1	BSII#1	Contig	1057	22.56528
GCF_004195735.1	DSM 40455	Contig	66	1.873297
GCF_004195745.1	NBRC 100770	Contig	41	1.571027
GCF_004195755.1	NBRC 13083	Contig	56	1.609157
GCF_004195765.1	CGMCC 4.1677	Contig	66	1.767082
GCF_004195775.1	DSM 40233	Contig	68	1.747961
GCF_004195835.1	FXJ8.011	Contig	48	1.830963
GCF_004195845.1	CGMCC 4.1681	Contig	54	1.92619
GCF_004195855.1	FXJ.2339	Contig	49	1.561473
GCF_004195885.1	NBRC 12790	Contig	53	1.57254

GCF_004195895.1	Ma24	Contig	56	1.594031
GCF_004195935.1	CR33	Contig	61	1.909779
GCF_004195945.1	CR10	Contig	38	1.734104
GCF_004195955.1	CR25	Contig	47	1.758422
GCF_004195965.1	Ma1	Contig	36	1.688646
GCF_004195995.1	CR47	Contig	38	1.538716
GCF_004196035.1	CR19	Contig	44	1.703691
GCF_004196045.1	CR46	Contig	74	1.956849
GCF_004196055.1	CR16	Contig	146	3.565336
GCF_004196095.1	CR15	Contig	60	2.34654
GCF_004196115.1	FXJ8.008	Contig	147	5.511811
GCF_004196125.1	CR13	Contig	75	1.682935
GCF_004196135.1	D62	Contig	87	2.197802
GCF_004196145.1	FXJ8.031	Contig	39	1.527495
GCF_004196155.1	CGMCC 4.1615	Contig	39	1.509754
GCF_004196215.1	FXJ6.189	Contig	42	1.376721
GCF_007896905.1	77	Scaffold	100	2.949147
GCF_010550625.1	SID9939	Contig	613	16.32069
GCF_013693715.1	J1074/R2	Complete Genome	1	1.422282
GCF_015321285.1	ANT_B131	Contig	49	4.001356
GCF_015601545.1	W68	Complete Genome	1	3.373962
GCF_019219645.1	LGO-A3	Scaffold	19	1.243163
GCF_019219685.1	LGO-A19	Scaffold	22	1.191461
GCF_019286195.1	LGO-A16	Chromosome	8	1.220736
GCF_019286295.1	LGO-A23	Chromosome	24	1.439331
GCF_020521315.1	NBRC 12854	Contig	6	1.694358
GCF_023502165.1	220	Scaffold	175	3.091115
GCF_024592655.1	VIP_1	Contig	87	3.485952
GCF_025630915.1	Q	Contig	93	1.553792
GCF_026339675.1	NBC_01691	Contig	3	1.401944
GCF_026339695.1	NBC_01670	Contig	6	1.643041
GCF_026635925.1	J1074_D14	Complete Genome	1	1.082136
GCF_026636035.1	J1074_D14 4N24	Complete Genome	1	1.43986
GCF_030345275.1	MGMM6	Complete Genome	1	1.600545
GCF_034479045.1	ATCC 23899	Complete Genome	1	1.320132
GCF_035955795.1	NBC_01331	Complete Genome	2	1.525124
GCF_036238395.1	NBC_01271	Complete Genome	1	1.133097
GCF_036239155.1	NBC_01270	Complete Genome	1	1.175483
GCF_036247045.1	NBC_01722	Complete Genome	1	1.563025
GCF_038242915.1	RKJM-0023	Complete Genome	2	1.472048
GCF_038593545.1	JAR6	Contig	79	2.918354
GCF_038662715.1	SC-3	Complete Genome	1	1.974448
GCF_040297505.1	NPDC001745	Scaffold	80	1.993687
GCF_040300425.1	NPDC000360	Scaffold	115	2.135825
GCF_040300445.1	NPDC000394	Scaffold	75	1.834711
GCF_040302045.1	NPDC000023	Scaffold	83	2.20013
GCF_040302805.1	NPDC097099	Scaffold	294	5.187773

GCF 040303165.1	NPDC096061	Scaffold	95	1.914507
GCF 040303845.1	NPDC094052	Scaffold	81	1.879637
GCF 040606235.1	NPDC005450	Scaffold	86	1.765893
GCF 040606265.1	NPDC005441	Scaffold	83	1.903226
GCF 040606295.1	NPDC005446	Scaffold	243	2.760373
GCF 040609275.1	NPDC004694	Scaffold	36	1.130039
GCF 040609495.1	NPDC004660	Scaffold	68	1.393319
GCF 040613135.1	NPDC006311	Scaffold	1518	42.41826
GCF 040613195.1	NPDC006304	Scaffold	209	3.915516
GCF 040613215.1	NPDC006300	Contig	393	8.045384
GCF 040628955.1	NPDC006027	Scaffold	42	1.530864
GCF 040629865.1	NPDC005868	Scaffold	219	3.707536
GCF 040630285.1	NPDC005817	Scaffold	146	1.818491
GCF 040633095.1	NPDC013402	Scaffold	79	2.407679
GCF 040633215.1	NPDC013284	Scaffold	97	2.173208
GCF 040634315.1	NPDC007115	Scaffold	101	2.996445
GCF 040634375.1	NPDC007114	Contig	874	17.1856
GCF 040634755.1	NPDC007058	Scaffold	96	3.521717
GCF 040634955.1	NPDC007024	Scaffold	100	2.026359
GCF 040635305.1	NPDC006977	Scaffold	88	3.173795
GCF 040635445.1	NPDC006965	Scaffold	43	1.389348
GCF 040635525.1	NPDC006953	Scaffold	96	3.704313
GCF 040635865.1	NPDC006834	Scaffold	77	2.70783
GCF 040636535.1	NPDC036225	Scaffold	102	3.679101
GCF 040637815.1	NPDC028204	Scaffold	41	1.499001
GCF 040640855.1	NPDC021591	Scaffold	71	1.630887
GCF 040641535.1	NPDC020773	Scaffold	113	1.81388
GCF 040642355.1	NPDC047859	Scaffold	77	2.903719
GCF 040643655.1	NPDC047089	Scaffold	53	1.824699
GCF 040644275.1	NPDC046998	Scaffold	72	1.497504
GCF 040644335.1	NPDC046987	Scaffold	59	1.503759
GCF 040644815.1	NPDC007108	Scaffold	197	3.787506
GCF 040646775.1	NPDC045167	Scaffold	50	2.520718
GCF 040647175.1	NPDC044138	Scaffold	93	3.147624
GCF 040677505.1	NPDC053831	Scaffold	132	3.110048
GCF 040677795.1	NPDC053738	Scaffold	137	2.044264
GCF 040677835.1	NPDC053736	Scaffold	60	1.6515
GCF 040677855.1	NPDC053732	Scaffold	75	1.482345
GCF 040678335.1	NPDC052703	Contig	328	6.744574
GCF 040680315.1	NPDC052085	Scaffold	80	2.098805
GCF 040681755.1	NPDC051581	Scaffold	74	1.664195
GCF 040685675.1	NPDC088199	Scaffold	201	3.435859
GCF 040685755.1	NPDC088191	Scaffold	85	1.438849
GCF 040701785.1	NPDC090300	Scaffold	48	1.408451
GCF 040701885.1	NPDC088366	Scaffold	324	3.351667
GCF 040701985.1	NPDC088200	Scaffold	117	2.17975
GCF 040702625.1	NPDC078892	Scaffold	83	2.000984

GCF_040703575.1	NPDC007091	Scaffold	102	1.639887
GCF_040703655.1	NPDC007077	Scaffold	83	3.225265
GCF_040706455.1	NPDC047995	Scaffold	55	1.471318
GCF_041438585.1	NBC_01791	Complete Genome	3	1.370086
GCF_041438605.1	NBC_01790	Complete Genome	2	2.668515
GCF_041438665.1	NBC_01747	Complete Genome	2	1.387529
GCF_041438825.1	NBC_01692	Complete Genome	2	1.346965
GCF_041438845.1	NBC_01616	Complete Genome	2	1.433349
GCF_042270365.1	NPDC055905	Scaffold	214	3.536115
GCF_042270905.1	NPDC055925	Scaffold	326	3.181603
GCF_042447705.1	S20	Complete Genome	2	2.756588
GCF_044626895.1	J29 ori2	Contig	61	1.04712
GCF_045039795.1	NBC_01170	Complete Genome	1	1.460339
GCF_045039995.1	NBC_01374	Complete Genome	1	1.37496
GCF_045209585.1	NBC_01621	Complete Genome	2	1.334201
GCF_045209685.1	NBC_01627	Complete Genome	2	1.606294
GCF_045209765.1	NBC_01640	Complete Genome	2	1.854892
GCF_045210095.1	NBC_01658	Complete Genome	4	2.746511
GCF_045210185.1	NBC_01661	Complete Genome	2	1.784566
GCF_045276955.1	NBC_01665	Complete Genome	3	1.862137
GCF_045276985.1	NBC_01671	Complete Genome	2	1.948052
GCF_045277005.1	NBC_01673	Complete Genome	1	1.112772
GCF_045277015.1	NBC_01675	Complete Genome	2	1.341843
GCF_045748235.1	A1	Contig	78	1.764207
GCF_047784865.1	ChSI01	Scaffold	35	2.142129
GCF_048277575.1	J25	Contig	64	1.284435
GCF_049489975.1	NBC_01664	Complete Genome	4	1.474595
GCF_049844815.1	DT179	Scaffold	115	1.314695
GCF_049847085.1	DT221	Scaffold	100	1.354289
GCF_049856095.1	MR-37	Scaffold	72	1.229577
GCF_050353295.1	AS2	Contig	97	2.268705
GCF_050409345.1	2A	Contig	2	1.343602
GCF_050409365.1	1B	Contig	4	1.480599
GCF_050409385.1	1A	Contig	2	1.344262
GCF_051148575.1	ML27	Complete Genome	2	1.515892
GCF_052474615.1	DB7_152	Complete Genome	3	2.133424
GCF_052806495.1	SID1138	Contig	490	9.329695
GCF_052806595.1	SID1124	Contig	537	10.69678
GCF_052806715.1	SID1105	Contig	783	17.13615
GCF_056247565.1	G1	Complete Genome	1	1.163744
GCF_056933435.1	MZ8	Scaffold	21	1.400382
GCF_057319795.1	HNIBRBA13268	Contig	133	3.262558
GCF_058244835.1	SL05	Contig	3	1.917404
GCF_059491255.1	STR-S001	Complete Genome	1	1.933748
GCF_900171555.1	R-53649	Scaffold	377	7.998621
GCF_947311705.1	CCOS 2040	Complete Genome	3	5.278442

*The percentage of incomplete genes classified by the PGAP2 pipeline during preprocessing.

Table S3. ANI-defined genomic groups within the 151-genome *S. albidoflavus*-assigned cohort

Genomic group	n	ANI to NRRL B-1271 (%)	Within-group ANI (%)	Reference association
Group 1	125	97.644–100	97.507–99.999	Contains NRRL B-1271 and MUM256
Group 2	14	96.189–96.304	97.334–99.989	Divergent from NRRL B-1271-associated group
Group 3	7	95.910–96.065	99.153–99.743	Divergent from NRRL B-1271-associated group
Group 4	3	95.776–95.933	96.751–98.529	Divergent from NRRL B-1271-associated group
Group 5	2	95.634–95.987	97.918	Divergent from NRRL B-1271-associated group

ANI ranges were calculated from the 151-genome FastANI matrix. ANI to NRRL B-1271 represents the range observed for members of each genomic group relative to *S. albidoflavus* NRRL B-1271. Groups are used as descriptive genomic partitions and do not constitute formal species assignments.

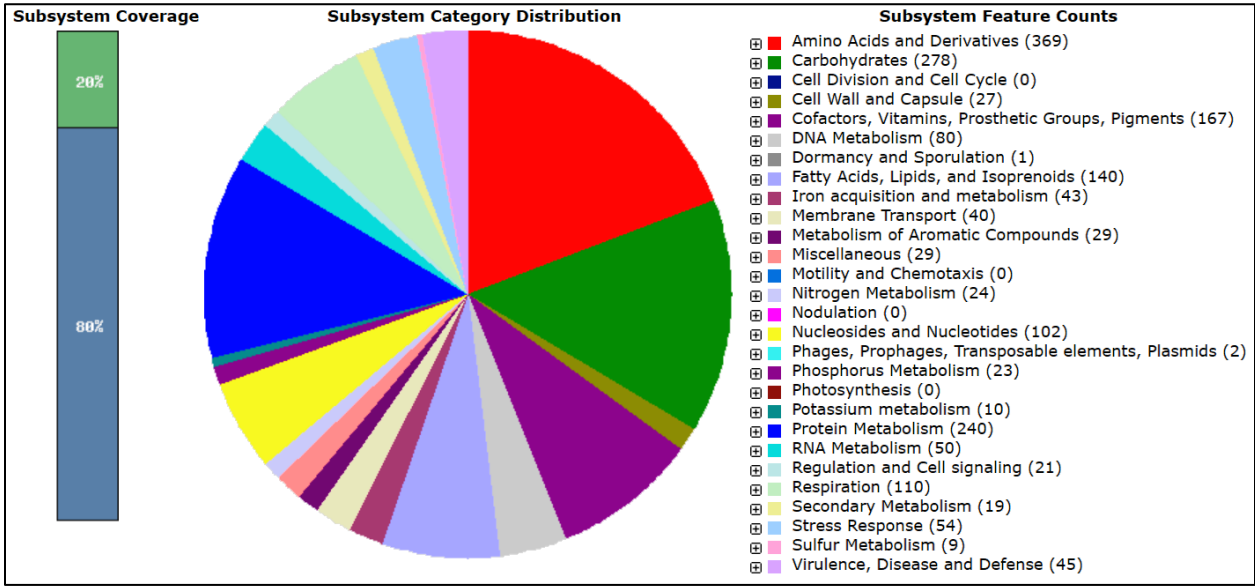


Figure S1. RAST subsystem assignments of *Streptomyces albidoflavus* MUM256.

A Cohort curation

Exclusions were predefined by study design and PGAP2 preprocessing QC.

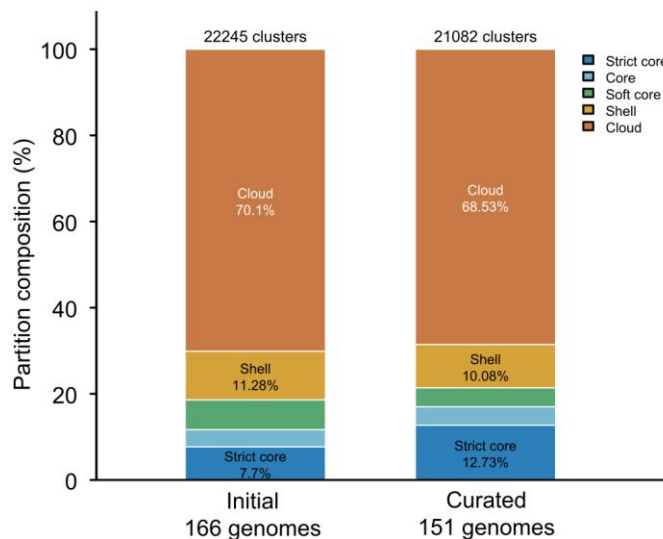
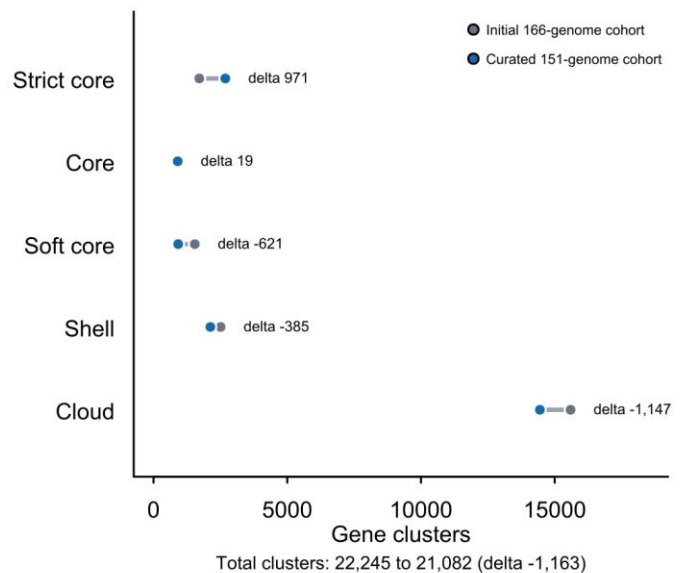
B Partition composition**C Absolute partition differences**

Figure S3. Cohort curation and sensitivity of PGAP2 pangenome partition structure. (A) Cohort-selection workflow. The initial 166-genome PGAP2 cohort was reduced by excluding three engineered strains (166 to 163 genomes) and 12 genomes with >10% incomplete/discarded PGAP2 genes, yielding the curated 151-genome cohort. These exclusions were predefined by study design and PGAP2 preprocessing quality control. (B) PGAP2 partition composition for the initial and curated cohorts. The five retained partition classes are strict core, core, soft core, shell and cloud. The initial cohort contains 1,713 strict-core, 885 core, 1,543 soft-core, 2,510 shell and 15,594 cloud clusters (22,245 total); the curated cohort contains 2,684, 904, 922, 2,125 and 14,447 clusters, respectively (21,082 total). Percentages are calculated within each cohort. (C) Absolute partition differences between cohorts. Points show cluster counts for the initial 166-genome cohort and curated 151-genome cohort; connecting segments show the change calculated as curated minus initial: +971 strict-core, +19 core, -621 soft-core, -385 shell and -1,147 cloud clusters (-1,163 clusters overall).

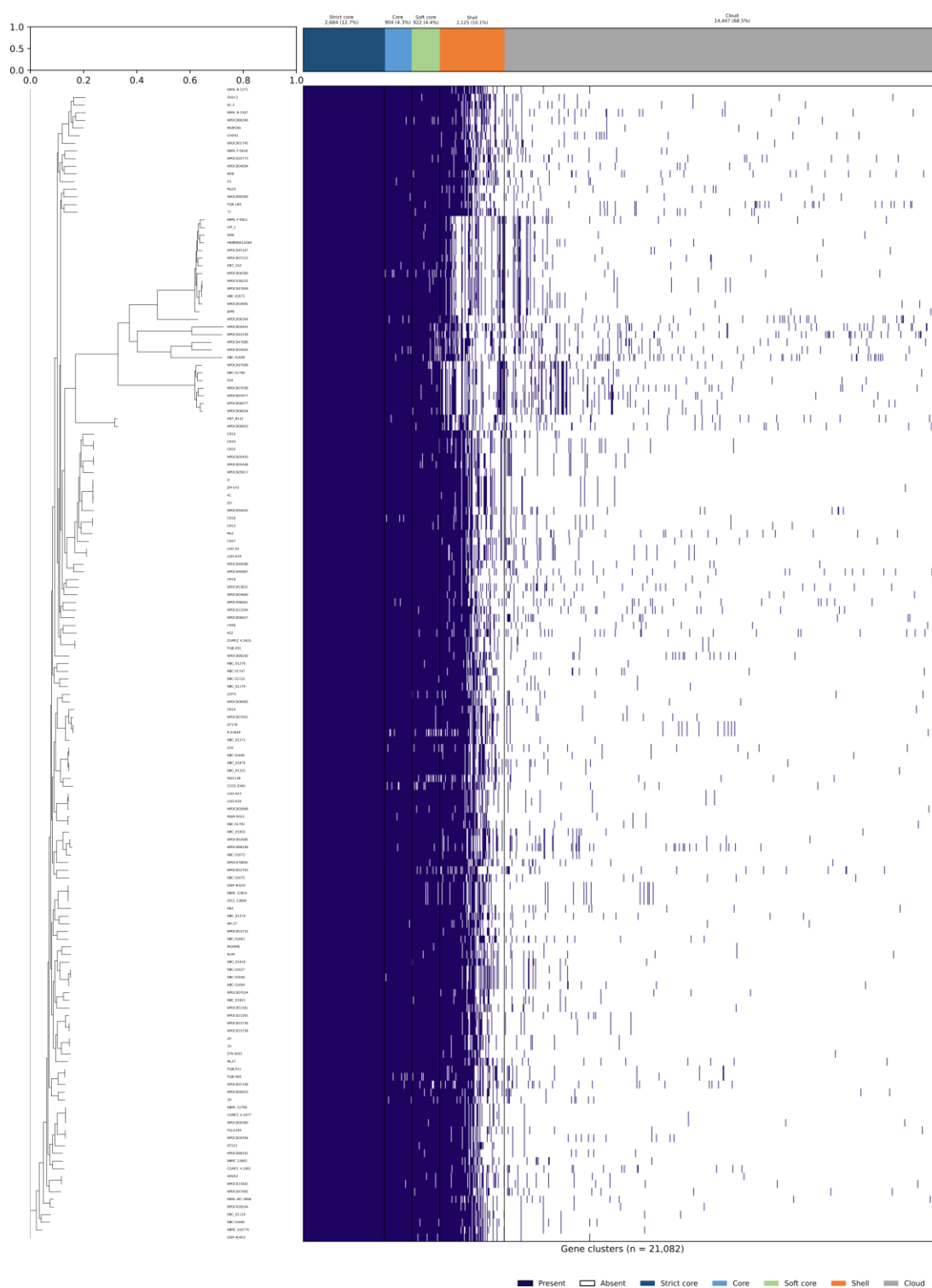


Figure S4. Gene-cluster presence/absence landscape across the 151-genome *Streptomyces albidoflavus* cohort. Presence/absence matrix of the 21,082 pangenome gene clusters identified by PGAP2 across 151 quality-filtered genomes. Rows represent individual genomes and columns represent gene clusters; dark blue cells denote presence and white cells

denote absence. Genomes are hierarchically arranged according to their gene-content profiles. For visualisation, gene clusters are grouped into core-associated (strict core, core, and soft core combined; 4,510 clusters, 21.4%), shell (2,125; 10.1%), and cloud (14,447; 68.5%) fractions, as indicated by the upper annotation bar. The matrix illustrates the broadly conserved core-associated component together with extensive strain-variable and sparsely distributed accessory gene content.

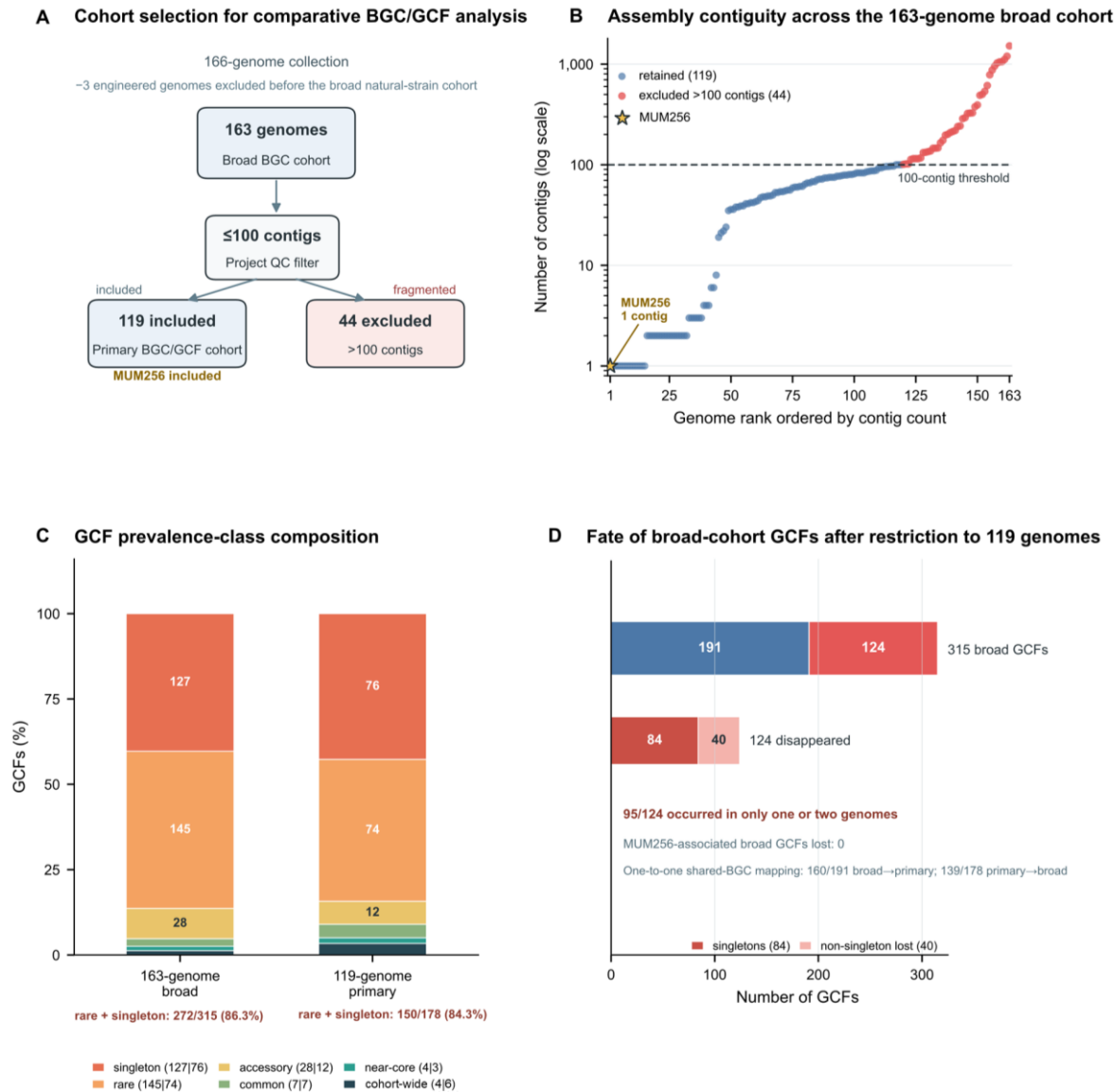


Figure S5. Cohort selection and sensitivity of population BGC/GCF structure to assembly-contiguity filtering. (A) Selection workflow from the 166-genome collection: three engineered genomes were excluded before formation of the 163-genome broad natural-strain BGC cohort, and 44 genomes with >100 contigs were excluded from the primary cohort,

leaving 119 genomes including MUM256. The ≤ 100 -contig criterion is a project-specific comparability safeguard, not a universal genome-quality threshold. (B) Ranked contig-count distribution for the 163-genome broad cohort; the dashed line marks 100 contigs. (C) Harmonised prevalence-class composition of run-specific GCFs in the broad 163-genome and primary 119-genome BiG-SCAPE analyses. Rare plus singleton GCFs accounted for 272/315 (86.3%) and 150/178 (84.3%), respectively. (D) Broad-cohort GCF fate after restriction to the primary cohort. Of 315 broad-cohort GCFs, 191 retained at least one BGC record and 124 disappeared; 84 of the disappearing families were singletons and 95 occurred in only one or two broad-cohort genomes. None of the disappearing families contained a MUM256 BGC. Cross-run membership matching showed one-to-one mapping for 160/191 retained broad families to a primary family and for 139/178 primary families to a broad family. Family labels are run-specific; these summaries do not imply that GCF identities are invariant across independent clustering runs.

region014 — cyclofaulknamycin-like BGC FAM_00112 | 44/119 | accessory | KnownClusterBlast similarity 100%

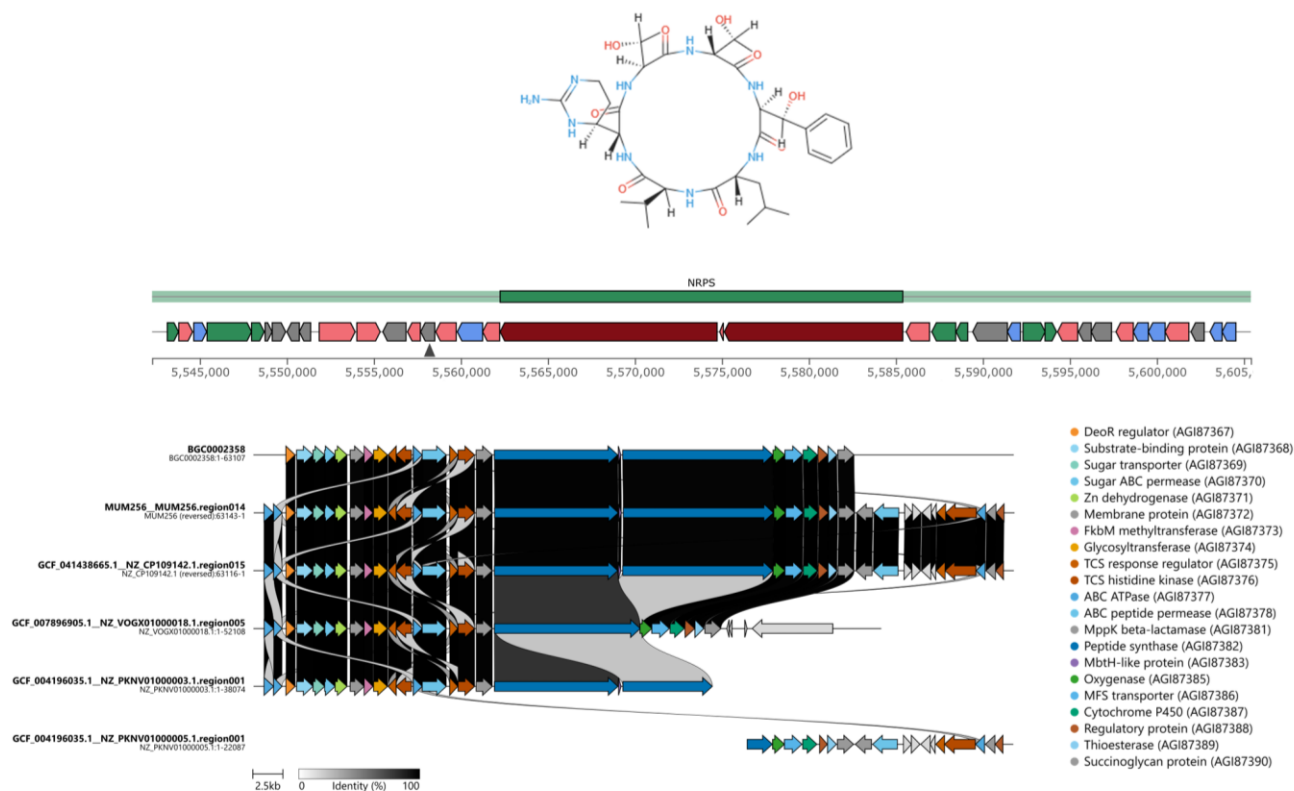


Figure S6. Comparative architecture of accessory cyclofaulknamycin-like biosynthetic loci in *Streptomyces albidoflavus* MUM256. MUM256 region014, assigned to FAM_00112 and present in 44 of 119 genomes (37.0%; accessory), showing its genomic context and comparative organisation relative to the characterised cyclofaulknamycin BGC BGC0002358.3 and representative FAM_00112 loci. antiSMASH KnownClusterBlast reported 100% similarity to the reference cluster.

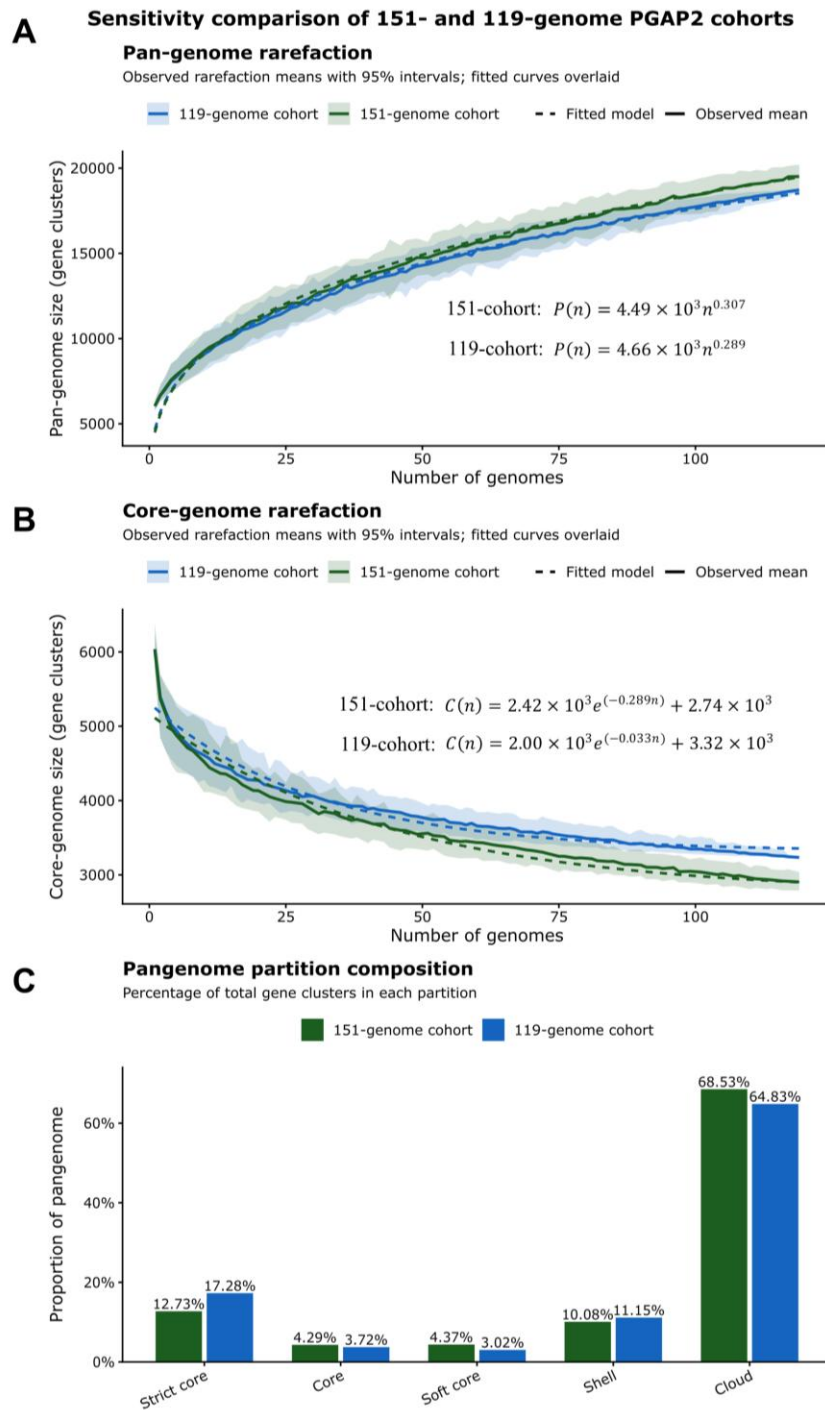


Figure S7. Sensitivity analysis of pangenome structure in the 151- and 119-genome PGAP2 cohorts. (A) Pan-genome rarefaction profiles for the 151-genome PGAP2 cohort and the more stringent 119-genome contiguity-filtered cohort. (B) Core-genome rarefaction profiles for the same cohorts, shown as observed mean trajectories with 95% resampling intervals and fitted exponential-asymptotic models. (C) Comparison of pangenome partition composition between the two cohorts. Strict-core, core, soft-core, shell and cloud clusters in the 151-genome cohort compared with the 119-genome cohort.