

University of Groningen

The Sequence of a 1.8-Mb Bacterial Linear Plasmid Reveals a Rich Evolutionary Reservoir of Secondary Metabolic Pathways

Medema, Marnix H.; Trefzer, Axel; Kovalchuk, Andriy; van den Berg, Marco; Muller, Ulrike; Heijne, Wilbert; Wu, Liang; Alam, Mohammad T.; Ronning, Catherine M.; Nierman, William C.

Published in:
Genome Biology and Evolution

DOI:
[10.1093/gbe/evq013](https://doi.org/10.1093/gbe/evq013)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2010

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Medema, M. H., Trefzer, A., Kovalchuk, A., van den Berg, M., Muller, U., Heijne, W., Wu, L., Alam, M. T., Ronning, C. M., Nierman, W. C., Bovenberg, R. A. L., Breitling, R., & Takano, E. (2010). The Sequence of a 1.8-Mb Bacterial Linear Plasmid Reveals a Rich Evolutionary Reservoir of Secondary Metabolic Pathways. *Genome Biology and Evolution*, 2, 212-224. <https://doi.org/10.1093/gbe/evq013>

Copyright

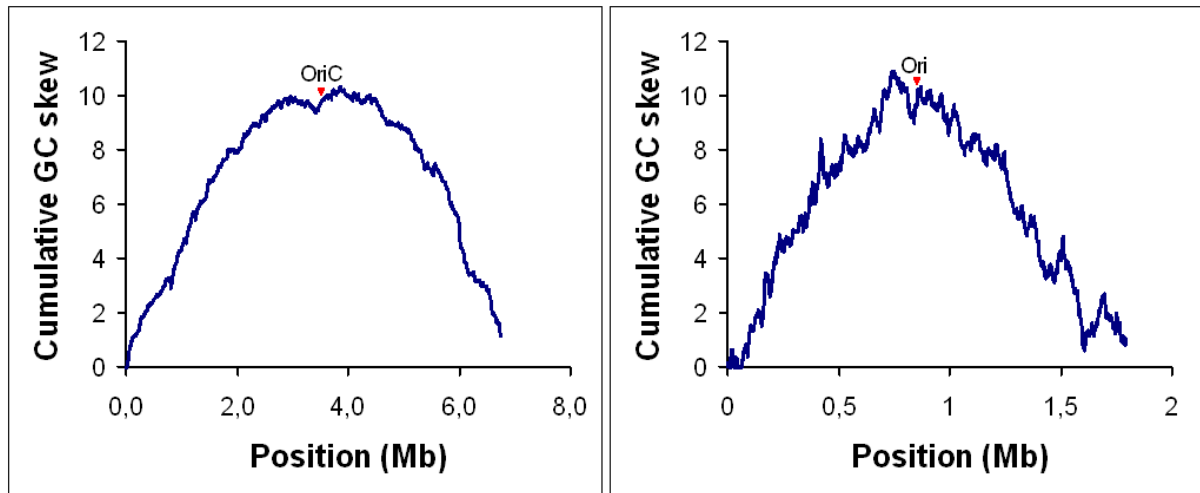
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

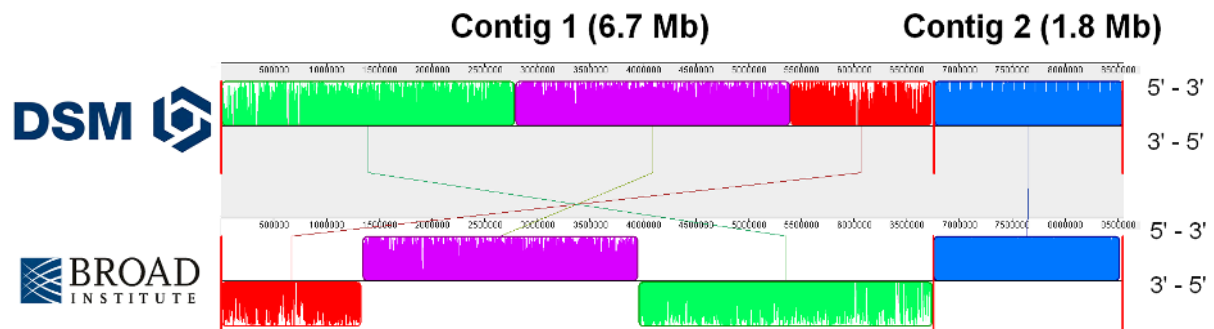
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



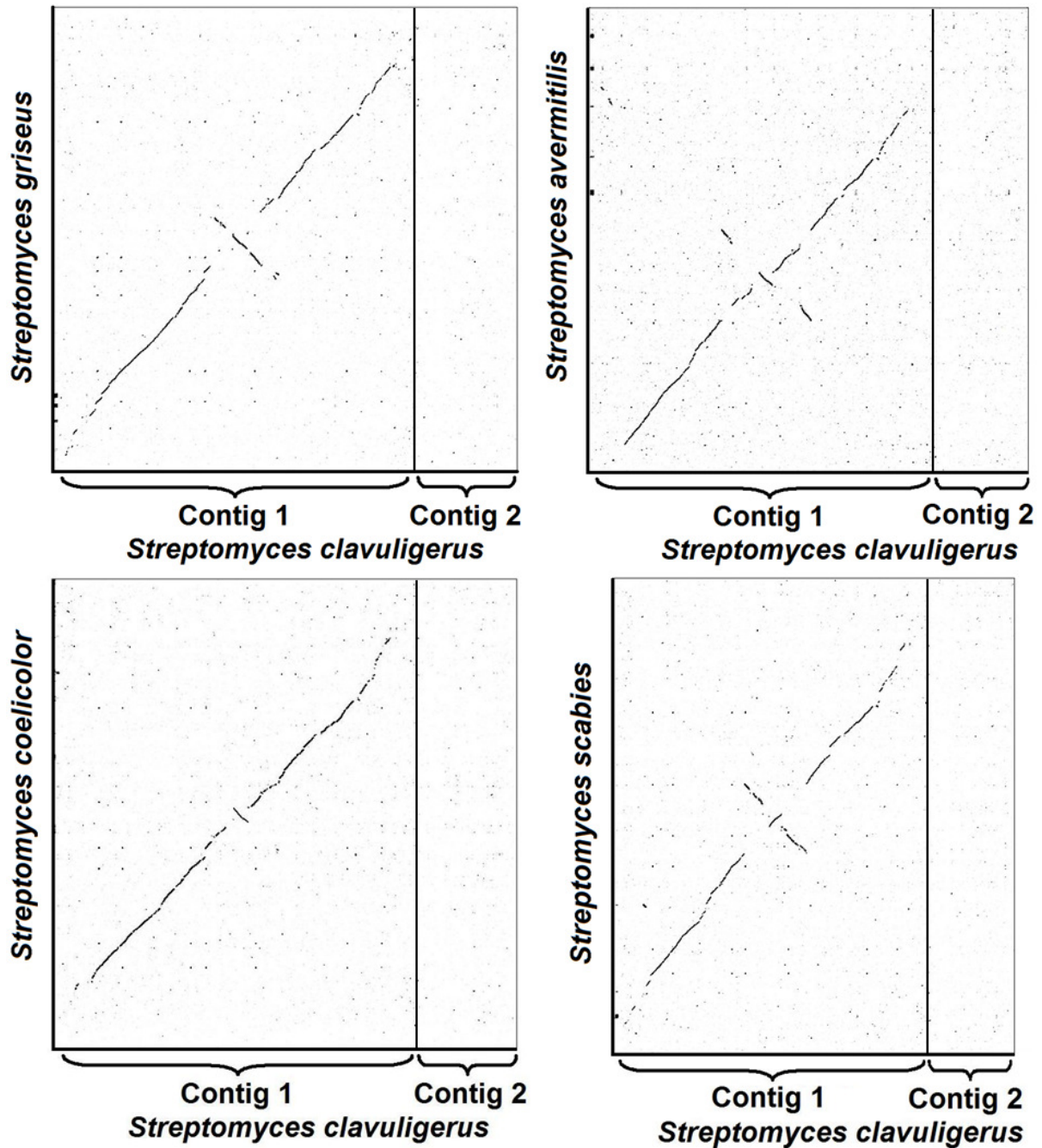
Supplementary Figure S1

Cumulative GC skew diagrams of the two contigs identified in the *S. clavuligerus* genome assembly. The position of the predicted origins of replication are shown as a red triangle. GC skew was calculated using GenSkew (<http://mips.gsf.de/services/analysis/genskew/>).



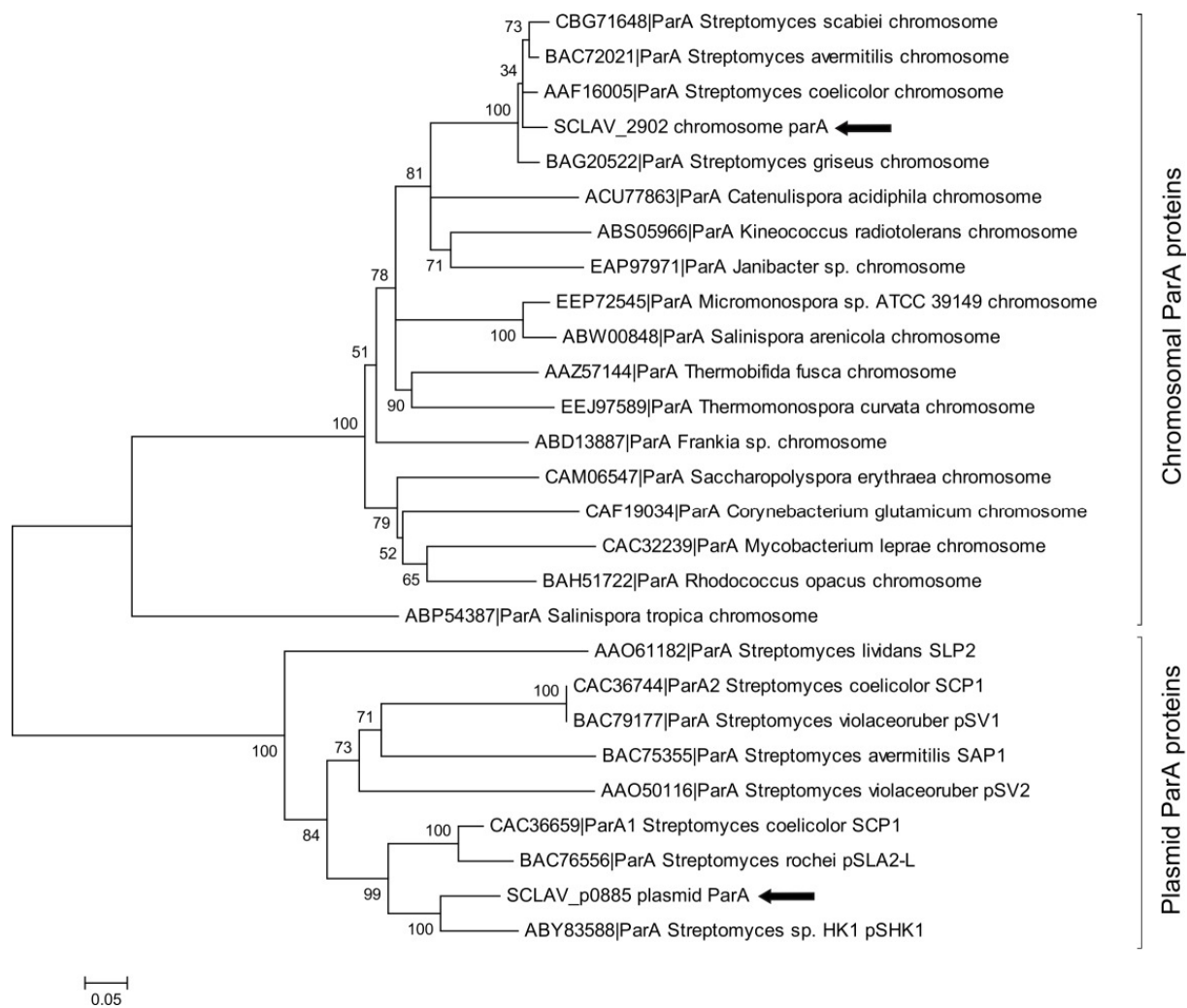
Supplementary Figure S2

Mauve (Darling et al. 2004) genome alignment of the DSM/JCVI and Broad Institute *S. clavuligerus* draft genome sequences. A large inversion of over 2.5 Mb is visible on the chromosome.



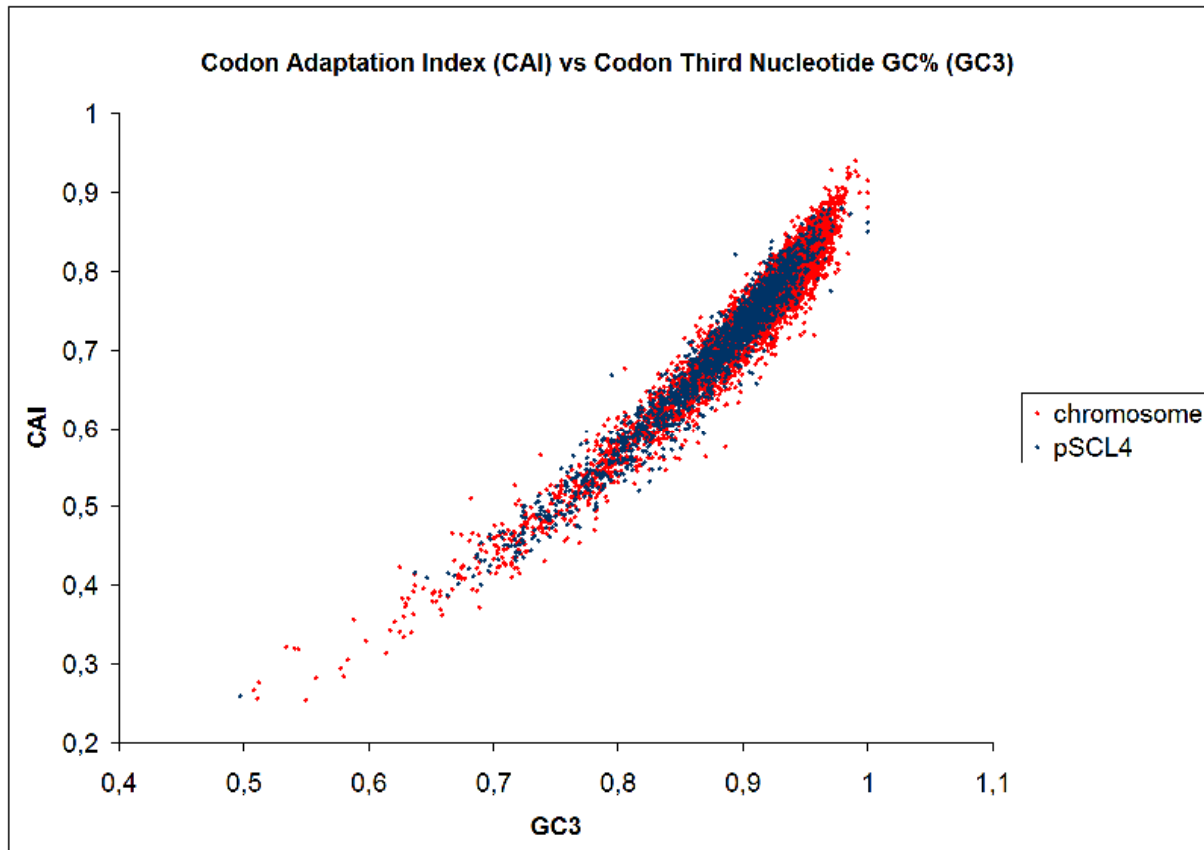
Supplementary Figure S3

Dot-plot comparisons of the *S. clavuligerus* genome with the four *Streptomyces* genomes completed so far. The dot-plots were constructed using Gepard (Krumsiek et al. 2007) on the complete nucleotide sequences of the composite genome of *S. clavuligerus*, and of the chromosomes of *S. coelicolor* A3(2), *S. griseus* IFO 13350, *S. avermitilis* ATCC 31267 and *S. scabies* 87.22.



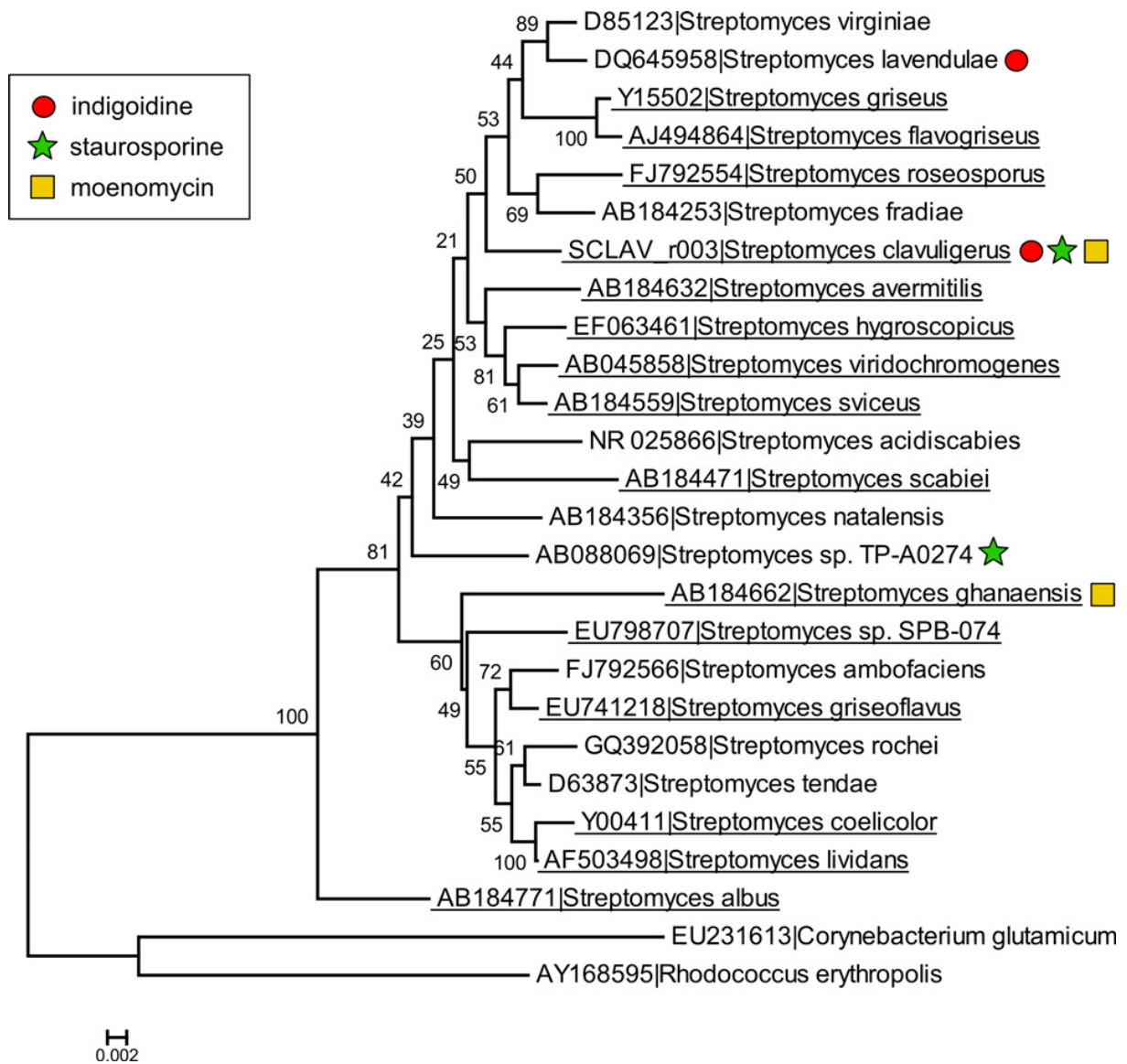
Supplementary Figure S4

Phylogeny of ParA proteins, which shows the distinctiveness of the plasmid and chromosomal maintenance machineries. Phylogenies were calculated using the NJ method in MEGA 4 (Tamura et al. 2007). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches.



Supplementary Figure S5

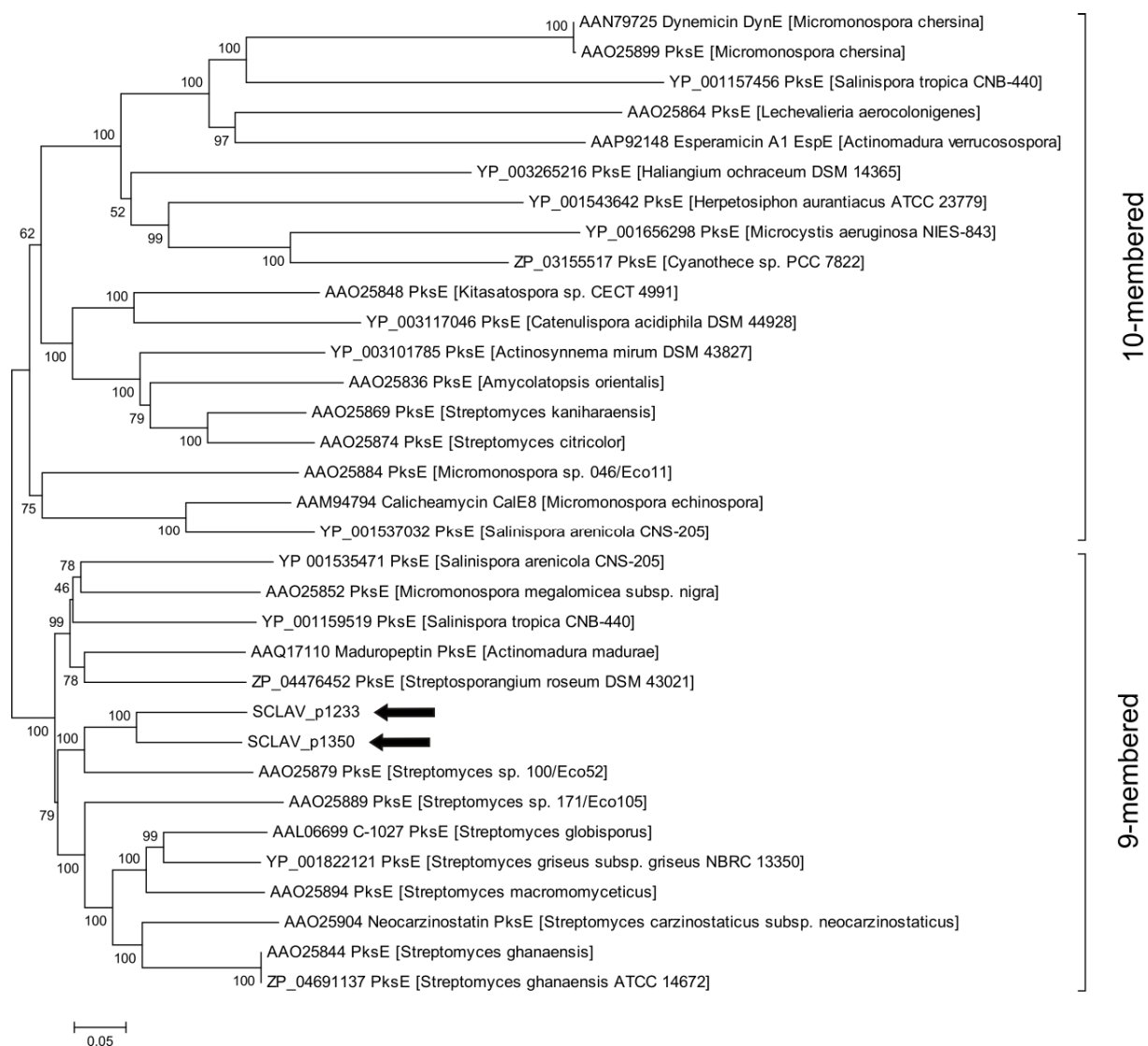
Distribution of all *S. clavuligerus* proteins from the chromosome and plasmid as to their Codon Adaptation Index (CAI) and third codon position GC % (GC3). CAI and GC3 were calculated with CAIcal (Puigbo et al. 2008) and CodonW (<http://codonw.sourceforge.net>).



Supplementary Figure S6

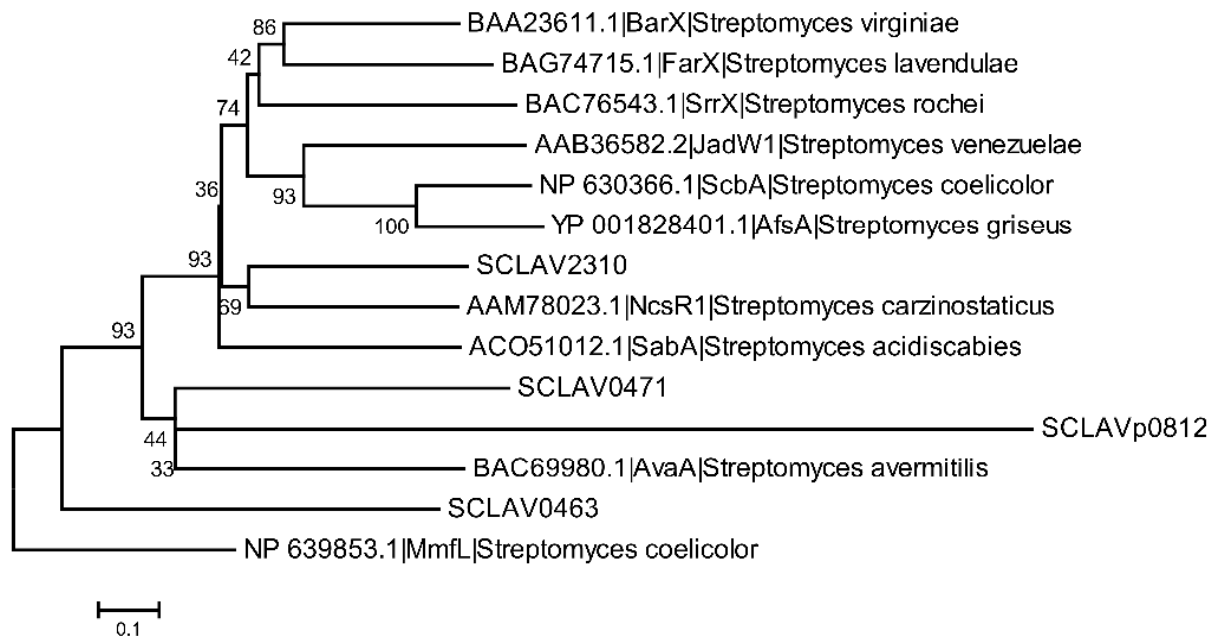
16S ribosomal RNA-based phylogenetic tree of *Streptomyces* species, showing the distribution of known moenomycin (yellow rectangle), staurosporine (green star) and indigoidine (red circle) gene clusters. The names of species of which either a complete genome or an almost complete draft genome is available are underlined. The dispersed absence/presence pattern indicates a probable role of horizontal gene transfer in the evolutionary history of these gene clusters. Phylogenies were calculated using the minimum evolution method in MEGA 4 (Tamura et al. 2007). The percentage of replicate trees in which

the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches.



Supplementary Figure S7

Phylogeny of enediyne-type PKSs, showing the position of the two enediyne-type PKSs detected on the megaplasmid pSCL4. Phylogenies were calculated using the neighbour-joining method in MEGA 4 (Tamura et al. 2007). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches.



Supplementary Figure S8

Phylogeny of AfsA-domain containing proteins, showing the clustering of SCLAV_2310 with the known γ -butyrolactone biosynthesis proteins ScbA and AfsA and their known homologs BarX, FarX, SrrX, JafW1, NcsR1 and SabA. The Mm furan biosynthesis protein MmfL from *S. coelicolor* (Corre et al. 2008) was used as an outgroup. Phylogenies were calculated using the neighbour-joining method in MEGA 4 (Tamura et al. 2007). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches.

Query (from <i>S. griseus</i>)	Hit	Identity (%)	Annotation	Unique to plasmid?
BAG20055.1 putative_cold_shock_protein	SCLAV_p0115	62.9	Cold-shock domain family protein	no
BAG20062.1 putative_ABC-transporter_ATP-binding_protein	SCLAV_p0477	71.29	ABC transporter ATP-binding protein	yes
BAG19908.1 putative_transcriptional_regulator	SCLAV_p0530	69.04	Transcriptional regulator	yes
BAG19876.1 putative_TetR-family_transcriptional_regulator	SCLAV_p0668	63.04	Putative TetR-family transcriptional regulator	yes
BAG20858.1 putative_ABC_transporter_ATP-binding_protein	SCLAV_p0830	60.79	Daunorubicin resistance ABC transporter ATP-binding subunit	yes
BAG20505.1 conserved_hypothetical_protein	SCLAV_p0866	66.76	Methicillin resistance protein	no
BAG21602.1 putative_UvrA-like_protein	SCLAV_p0905	80.7	UvrABC system protein A	no
BAG18126.1 putative_malate_synthase	SCLAV_p0928	61.44	Malate synthase	no
BAG18044.1 putative_NAD-dependent_aldehyde_dehydrogenase	SCLAV_p0939	86.17	Methylmalonate-semialdehyde dehydrogenase	yes
BAG18460.1 putative_monooxygenase	SCLAV_p0942	66.67	2-polyprenyl-6-methoxyphenol hydroxylase-like oxidoreductase	yes
BAG17193.1 putative_methyltransferase	SCLAV_p1166	74.27	Putative methyltransferase	yes
BAG19507.1 putative_cation_transport_system_component	SCLAV_p1322	62.13	Potassium-transporting ATPase subunit B	no
BAG20340.1 putative_ATP-dependent_RNA_helicase	SCLAV_p1485	62.79	ATP-dependent RNA helicase	no
BAG22539.1 putative_iron-siderophore_uptake_ABC_transporter_ATP-binding_protein	SCLAV_p1592	67.33	Ferric enterobactin transport ATP-binding protein FepC	no

Supplementary Table S1

Core genome hits on the 1.8 Mb replicon.

Name	Genes	Type	Putative function / remarks
SMCp1	SCLAV_p0064 – SCLAV_p0075	Terpene synthase	
SMCp2	SCLAV_p0286 – SCLAV_p0308	NRPS & PKS-NRPS	
SMCp3	SCLAV_p0326 – SCLAV_p0334	Terpene synthase & Type II PKS	Pentalenene synthase
SMCp4	SCLAV_p0486 – SCLAV_p0492	Terpene synthase	
SMCp5	SCLAV_p0509 – SCLAV_p0520	Type II PKS	
SMCp6	SCLAV_p0563 – SCLAV_p0588	NRPS / PKS / Terpene synthase	TS = Pentalenene synthase
SMCp7	SCLAV_p0633 – SCLAV_p0643	Terpene synthase	Germacradienol/germacrene D synthase
SMCp8	SCLAV_p0690 – SCLAV_p0714	Other	
SMCp9	SCLAV_p0760 – SCLAV_p0768	Terpene synthase	Pentalenene synthase
SMCp10	SCLAV_p0966 – SCLAV_p0986	Terpene synthase	Pentalenene synthase, terpene cyclase
SMCp11	SCLAV_p1005 – SCLAV_p1012	PKS-NRPS	
SMCp12	SCLAV_p1057 – SCLAV_p1063	Actin-ADP-ribosylating toxin	
SMCp13	SCLAV_p1074 – SCLAV_p1079	Beta-lactam	'Paralogous' alanylclavam cluster
SMCp14	SCLAV_p1107 – SCLAV_p1123	Indolocarbazole	Staurosporine gene cluster
SMCp15	SCLAV_p1170 – SCLAV_p1194	Terpene synthase	Pentalenene synthase (2x)
SMCp16	SCLAV_p1219 – SCLAV_p1234	Enediyne PKS	
SMCp17	SCLAV_p1263 – SCLAV_p1273	PKS-NRPS	
SMCp18	SCLAV_p1274 – SCLAV_p1290	Phosphoglycolipid	Moenomycin gene cluster
SMCp19	SCLAV_p1302 – SCLAV_p1311	NRPS	
SMCp20	SCLAV_p1334 – SCLAV_p1341	NRPS	NRPS fused to major facilitator transporter in final module
SMCp21	SCLAV_p1345 – SCLAV_p1376	Enediyne PKS	
SMCp22	SCLAV_p1407 – SCLAV_p1415	Terpene synthase	Pentalenene synthase
SMCp23	SCLAV_p1426 – SCLAV_p1432	NRPS / Terpene synthase	
SMCp24	SCLAV_p1471 – SCLAV_p1483	NRPS	Indigoidine synthetase
SMCp25	SCLAV_p1508 – SCLAV_p1510	Possible beta-lactam cluster	
SMC1	SCLAV_0001 – SCLAV_0026	Macrolide type I PKS	
SMC2	SCLAV_0082 – SCLAV_0105	Other	Putative secondary metabolite biosynthesis cluster without known homologs
SMC3	SCLAV_0148 – SCLAV_0152	Siderophore	
SMC4	SCLAV_0153 – SCLAV_0172	Terpene synthase	Germacradienol synthase-like
SMC5	SCLAV_0446 – SCLAV_0497	Type I PKS	
SMC6	SCLAV_1942 – SCLAV_1955	Siderophore	Desferrioxamines cluster
SMC7	SCLAV_2274 – SCLAV_2302	NRPS	
SMC8	SCLAV_2456 – SCLAV_2469	Lantibiotic	
SMC9	SCLAV_2920 – SCLAV_2935	Beta-lactam	Clavams
SMC10	SCLAV_4178 – SCLAV_4197	Beta-lactam	Clavulanic Acid
SMC11	SCLAV_4198 – SCLAV_4217	NRPS / Beta-lactam	Cefamycin C
SMC12	SCLAV_4387 – SCLAV_4392	Lantibiotic	
SMC13	SCLAV_4460 – SCLAV_4486	PKS-NRPS	
SMC14	SCLAV_4740 – SCLAV_4758	NRPS	NRPS-Beta-lactamase fusion in one module
SMC15	SCLAV_4911 – SCLAV_4921	Type II PKS	Putative spore pigment (WhiEI and WhiEII homologs)
SMC16	SCLAV_5142 – SCLAV_5148	NRPS	
SMC17	SCLAV_5227 – SCLAV_5255	Phytoene / squalene	
SMC18	SCLAV_5267 – SCLAV_5278	NRPS	
SMC19	SCLAV_5325 – SCLAV_5347	NRPS	
SMC20	SCLAV_5463 – SCLAV_5497	Type III PKS	
SMC21	SCLAV_5608 – SCLAV_5638	PKS-NRPS	
SMC22	SCLAV_5670 – SCLAV_5674	Terpene synthase	
SMC23	SCLAV_5702 – SCLAV_5710	Siderophore	IucA/IucC

Supplementary Table S2

Secondary metabolism gene clusters in the *S. clavuligerus* genome. SMC = secondary metabolite cluster.

SMC1-23 are the chromosome SMCs, SMCp1-p25 are the plasmid SMCs.

Gene	Annotation
SCLAV_0170	Beta-lactamase
SCLAV_1453	Alkaline D-peptidase
SCLAV_1622	Beta-lactamase
SCLAV_1657	Beta-lactamase
SCLAV_2072	D-alanyl-D-alanine carboxypeptidase
SCLAV_2296	Putative esterase
SCLAV_2436	Beta-lactamase
SCLAV_2579	Beta-lactamase
SCLAV_2874	Beta-lactamase
SCLAV_3000	D-alanyl-D-alanine carboxypeptidase
SCLAV_3128	Alkaline D-peptidase
SCLAV_3176	Alkaline D-peptidase
SCLAV_3177	Alkaline D-peptidase
SCLAV_4216	Beta-lactamase
SCLAV_4375	Beta-lactamase
SCLAV_4749	Non-ribosomal peptide synthetase with beta-lactamase domain
SCLAV_4822	Alkaline D-peptidase
SCLAV_p0031	Serine-type D-Ala-D-Ala carboxypeptidase
SCLAV_p0119	D-stereospecific peptide hydrolase precursor
SCLAV_p0171	Putative beta-lactamase
SCLAV_p0883	Penicillin-binding protein, beta-lactamase class C
SCLAV_p1311	Penicillin-binding protein, beta-lactamase class C

Supplementary Table S3

Beta-lactamase domain-containing proteins in the *S. clavuligerus* genome.

References

- Corre C, Song L, O'Rourke S, Chater KF, and Challis GL. 2008. 2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids, antibiotic production inducers discovered by *Streptomyces coelicolor* genome mining. *Proc Natl Acad Sci USA*. **105**:17510-17515.
- Darling AC, Mau B, Blattner FR, and Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*. **14**:1394-1403.
- Guindon S, Delsuc F, Dufayard JF, and Gascuel O. 2009. Estimating maximum likelihood phylogenies with PhyML. *Methods Mol Biol*. **537**:113-137.
- Krumsiek J, Arnold R, and Rattei T. 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics*. **23**:1026-1028.
- Puigbo P, Bravo IG, and Garcia-Vallve S. 2008. CAIcal: a combined set of tools to assess codon usage adaptation. *Biol Direct*. 3:38
- Tamura K, Dudley J, Nei M, and Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol*. **24**:1596-1599.