

## Type I Polyketide Synthases May Have Evolved Through Horizontal Gene Transfer

Aurélien Ginolhac,<sup>1,2</sup> Cyrille Jarrin,<sup>1,2</sup> Patrick Robe,<sup>1</sup> Guy Perrière,<sup>3</sup> Timothy M. Vogel,<sup>2</sup> Pascal Simonet,<sup>2</sup> Renaud Nalin<sup>1</sup>

<sup>1</sup> LibraGen S.A., Bâtiment Canal Biotech 1, 3 rue des Satellites, 31400 Toulouse, France

<sup>2</sup> Ecologie Microbienne, UMR CNRS 5557, Université Claude Bernard, Lyon 1, 43 boulevard du 11 Novembre 1918, 69622 Villeurbanne Cedex, France

<sup>3</sup> Laboratoire de Biométrie et Biologie Évolutive, UMR CNRS 5558, Université Claude Bernard, Lyon 1, 43 boulevard du 11 Novembre 1918, 69622 Villeurbanne Cedex, France

Received: 26 May 2004 / Accepted: 2 February 2005 [Reviewing Editor: Dr. Nicolas Galtier]

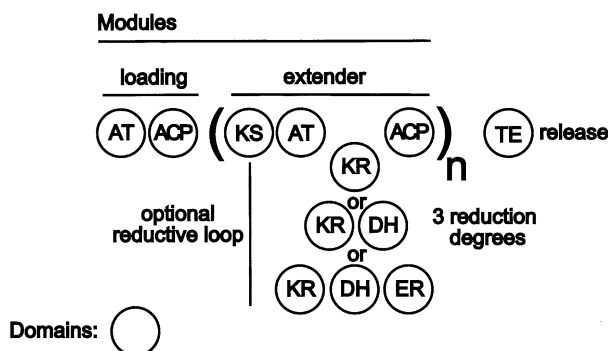
**Abstract.** Type I polyketide synthases (PKSI) are modular multidomain enzymes involved in the biosynthesis of many natural products of industrial interest. PKSI modules are minimally organized in three domains: ketosynthase (KS), acyltransferase (AT), and acyl carrier protein. The KS domain phylogeny of 23 PKSI clusters was determined. The results obtained suggest that many horizontal transfers of PKSI genes have occurred between actinomycetales species. Such gene transfers may explain the homogeneity and the robustness of the actinomycetales group since gene transfers between closely related species could mimic patterns generated by vertical inheritance. We suggest that the linearity and instability of actinomycetales chromosomes associated with their large quantity of genetic mobile elements have favored such horizontal gene transfers.

**Key words:** Polyketide synthase — Horizontal gene transfer — Molecular phylogeny — Ketosynthase — Actinomycetales

### Introduction

Polyketide synthases type I (PKSI) are large enzymes that are responsible for the biosynthesis of polyketides, natural products of pharmaceutical and industrial interest such as erythromycin (antibiotic), rapamycin (immunosuppressant), and sorafen (antifungal). Among the three types of bacterial PKS that have been reported to date (Shen 2003), only type I shows a modular organization, allowing the comparative study of modules within or between enzymes. The PKSI consists of multifunctional enzymes grouped in modules (Fig. 1). A biosynthesis pathway is commonly composed of several clustered PKSI genes. Proteins encoded by these clustered PKSI genes catalyze the linear steps of polyketide biosynthesis. As an example, three PKSI proteins, called eryA, eryB, and eryC, (Fig. 2) required for the biosynthesis of erythromycin. Each PKSI protein loads and condenses two acyl groups to build the erythromycin (Staunton and Weissman 2001). The three required PKSI proteins are called “erythromycin synthase.” In this study, the term “X synthase” refers to any of the PKSI proteins involved in the synthesis of polyketide X.

Two types of module exist inside a PKSI protein: the first module is called the loading module, while the following ones are called extender modules (Fig. 1). Each extender module receives the incoming acyl chain from the previous module and catalyzes the



**Fig. 1.** General representation of the PKS I organization. The PKS I are usually composed of a loading module followed by  $n$  extension modules building the growing acyl chain. Each substrate can be reduced by an optional reductive loop from one to three steps. A last domain thioesterase, is often present and catalyzes the release of the final product. KS, ketosynthase; AT, acyltransferase; ACP, acyl carrier protein; DH, dehydratase; KR, ketoreductase; ER, enoylreductase and TE, thioesterase.

condensation of the acyl groups to the growing acyl chain. Then the module transfers the acyl chain to the next module for an additional elongation step (Watanabe et al. 2003). A minimal extender module consists of three functional domains: a ketosynthase domain (KS), an acyltransferase domain (AT), and an acyl carrier protein (ACP). The KS domain catalyzes the decarboxylative condensation of an acyl group onto the growing chain. The AT domain selects and activates the growing chain. The ACP domain loads the growing chain (Staunton and Weissman 2001). The presence of additional domains with enzymatic activities of ketoreductase (KR), dehydratase (DH), and enoylreductase (ER) can modify the growing acyl chain (Hopwood 1997). Usually, the last domain of PKS I is a thioesterase, which catalyzes the release and the cyclization of the completed acyl chain (Fig. 1).

Several synthases combine both PKS I and non-ribosomal peptide synthases (NRPS) modules. The NRPS are also modular multienzymes, which build peptides with amino acids as the PKS I build polyketides with acyl groups (Du et al. 2001).

Hopwood (1997) proposed that the PKS I diversity has evolved by gene duplication followed by sequence divergence. In addition to this mechanism, horizontal gene transfer (HGT) has been invoked as a major force in bacterial evolution (Gogarten et al. 2002; Lopez 2003). Even if HGT must be invoked with precaution (Kurland et al. 2003), the specific genetic context of secondary metabolism in actinomycetales supports the view of many HGTs within this order. The complete genome sequence of *Streptomyces coelicolor* reveals 23 gene clusters devoted to secondary metabolism and representing a significant fraction (4.5%) of the genome (Challis and Hopwood 2003). The extensive genetic diversity of biosynthetic

clusters among actinomycetales (especially that of genus *Streptomyces*) exceeds a theoretical phenotypic potential and, therefore, supports HGT (Weber et al. 2003). The aim of this paper is to highlight the PKS I gene's evolutionary history within bacteria using phylogenetic reconstructions from well-characterized PKS I clusters.

## Materials and Methods

### Published Sequences

Polyketide synthases were selected when the produced polyketides were experimentally characterized. Thus, 23 biosynthetic pathways encompassing 203 protein sequences of the KS domain were computed (Fig. 2). The KS domain acts as a generic condensing function, whereas the AT domain is substrate-specific. The KS domain was retained for phylogenetic analyses. The AT and ACP domains were analyzed in the particular case of the close relationship among amphotericin, nystatin, and pimaricin synthases. The ACP domain holds the growing chain whatever its length. This domain would have been useful since, like the KS domain, its function is generic. However, the ACP domain does not conceal enough informative sites due to its short length. Thus, the phylogeny of ACP domains was determined only for the closest polyketide synthases.

### Phylogenetic Reconstruction and Analysis

Domains of public protein sequences were aligned using DbClustal (Thompson et al. 2000). Then the alignments were manually corrected using SeaView (Galtier et al. 1996). Phylogenetic reconstructions were performed via the distance method using neighbor joining (NJ) with the PAM matrix through Phylo\_win (Galtier et al. 1996) and the maximum likelihood method (ML) using the Jones–Taylor–Thornton (JTT; 1992) model of substitution through PhyML (Guindon and Gascuel 2003). All trees were built with 500 bootstrap replicates. For the ML reconstruction, 500 data sets were generated using the SEQBOOT program from the PHYLIP package v3.6 (Felsenstein 1993). A tree was built for each replicate with PhyML, then bootstraps were computed with CONSENSE (Felsenstein 1993). All trees were drawn with NJplot (Perrière and Gouy 1996) and rooted with a fatty acid synthase (mas; Uniprot accession No. Q02251), which is a protein family that has evolved apart from PKS I (Hopwood 1997; Zhu et al. 2002).

Relative rate tests were performed with RRTree (Robinson-Rechavi and Huchon 2000). This software compares substitution rates between DNA and protein sequences grouped in phylogenetically defined lineages. The Kishino–Hasegawa–Templeton (KHT) test (Kishino and Hasegawa 1989) was computed using ProML (Felsenstein 1993). It allows testing of phylogenetic hypotheses through likelihood comparison of phylogenetic trees.

## Results

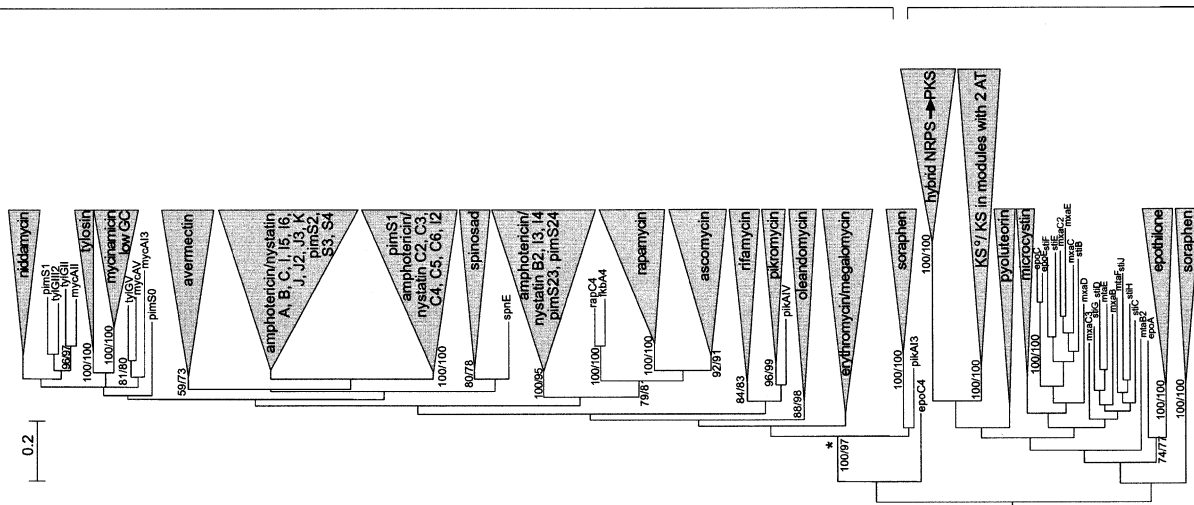
Three main arguments are assumed to underlie HGT: (i) exceptions to the clustering of KS domains from actinomycetales PKS I; (ii) the specific history of amphotericin, nystatin, and pimaricin synthases; and (iii) the dual origin of the soraphen synthase.

Organism (Phylum, order, species)	UniProt AC	Molecule	Protein name (Nb of modules)	Remarks	Reference
<b>Actinobacteria</b>					
<b>Actinomycetales</b>					
<i>Streptomyces nodosus</i>	Q93NW8	amphotericin	ampA (1)	KS <sup>+</sup>	Calfrey et al. 2001
	Q93NW7		ampB (2)		
	Q93NW6		ampC (6)		
	Q93NX9		ampI (6)		
	Q93NX8		ampJ (3)		
	Q93NX7		ampK (1)		
<i>Streptomyces hygroscopicus</i>	Q9KIE0	ascocycin	fkfB (4)		Wu et al. 2000
	Q9KIE1		fkfC (2)		
	Q9KID7		fkfA (4)		
<i>Streptomyces avermitilis</i>	Q9SOR8	avermectin	aveA1 (3)	loading: AT-ACP	Ikeda et al. 1999
	Q9SOR7		aveA2 (4)		
	Q9SOR4		aveA3 (3)		
	Q9SOR3		aveA4 (3)		
<i>Saccharopolyspora erythraea</i>	O03131	erythromycin	eryA (2)		Cortez et al. 1990
	O03132		eryB (2)		
	O03133		eryC (2)		
<i>Micromonospora megalomica</i>	Q9F830	megalomicin	megAI (2)		Vokhugursky et al. 2000
	Q9F829		megAll (2)		
	Q9F828		megAllI (2)		
<i>Micromonospora griseorubida</i>	Q83WF0	mycinamicin	mycAI (3)	KS <sup>+</sup> ; 2: low GC	Anzai et al. 2003
	Q83WE9		mycAll (1)		
	Q83WE8		mycAllI (2)		
	Q83WE7		mycAIV (1)		
<i>Streptomyces caelestis</i>	Q30764	niddamycin	nidA1 (3)	KS <sup>+</sup>	Kakavas et al. 1997
	Q30765		nidA2 (1)		
	Q30766		nidA3 (2)		
	Q30767		nidA4 (1)		
	Q30768		nidA5 (1)		
	Q9L4W5		nysA (1)		
<i>Streptomyces noursei</i>	Q9L4W4	nystatin	nysB (2)	KS <sup>+</sup>	Brautaset et al. 2000
	Q9L4W3		nysC (6)		
	Q9L4X3		nysI (6)		
	Q9L4X2		nysJ (3)		
	Q9L4X1		nysK (1)		
	Q9KIV4		oleA (2)		
<i>Streptomyces antibioticus</i>	Q9KIV3	oleandomycin	oleAll (2)	KS <sup>+</sup>	Shah et al. 2000
	Q07017		oleAllI (2)		
	Q9ZG15		pikAI (3)		
<i>Streptomyces venezuelae</i>	Q9ZG14	pikromycin	pikAll (2)	KS <sup>+</sup>	Xue et al. 2000
	Q9ZG13		pikAllI (1)		
	Q9ZG12		pikAIV (1)		
<i>Streptomyces natalensis</i>	Q9X992	pimamicin	pimS0 (1)	KS <sup>+</sup>	Aparicio et al. 2000
	Q9X993		pimS1 (4)		
	Q8EWA1		pimS2 (6)		
	Q8EWA2		pimS3 (1)		
	Q8EWA3		pimS4 (1)		
<i>Streptomyces hygroscopicus</i>	O54297	rapamycin	rapA (4)		Schwecke et al. 1995
	O54296		rapB (6)		
	O54299		rapC (4)		
	O54666		rilA (3)		
<i>Amycolatopsis mediterranei</i>	O52545	rilamycin	rilB (3)		Tang et al. 1998
	O52790		rilC (1)		
	O54591		rilD (1)		
	O54593		rilE (2)		
	Q9ALM6		spnA (2)		
<i>Saccharopolyspora spinosa</i>	Q9ALM5	spinosad	spnB (1)	KS <sup>+</sup>	Waldron et al. 2001
	Q9ALM4		spnC (2)		
	Q9ALM3		spnD (3)		
	Q9ALM2		spnE (3)		
	Q9ALM1		spnF (1)		
<i>Streptomyces fradiae</i>	O33954	tyloacin	tylGI (3)	KS <sup>+</sup>	Cundliffe et al. 2001
	O33955		tylGII (1)		
	O33956		tylGIII (2)		
	O33957		tylGIV (1)		
	O33958		tylGV (1)		
<b>Proteobacteria</b>					
<b>Pseudomonadales</b>					
<i>Pseudomonas fluorescens</i>	O51827	pyoluteorin	pIB (2)		Nowak-Thompson et al. 1997
	O51828		pIC (1)		
<b>Myxococcales</b>					
<i>Sorangium cellulosum</i>	Q9L8C9	epothione	epoA (1)	KS <sup>+</sup>	Molnar et al. 2000
	Q9L8C8		epoB (1)		
<i>Polyangium calidum</i>	Q9L8C7	soraphen	epoC (4)	preceded by NRPS	
	Q9L8C6		epoD (2)		
	Q9L8C5		epoE (1)		
	Q9ADL6		scrA (3)		
	O53840		scrB (5)		
	Q9RFLO		mtaB (2)		
<i>Stigmatella aurantiaca</i>	Q9RFK8	myxothiazol	mtaD (1)	loading: 2 AT preceded by NRPS	Slaskowski et al. 1999
	Q9RFK7		mtaE (1)		
	Q9RFK6		mtaF (1)		
	Q93TW6		mxsA (1)		
	Q93TW7		mxsE (1)		
<i>Stigmatella aurantiaca</i>	Q93TW8	myxalamid	mxsD (1)	loading: 2 AT	Slaskowski et al. 2001
	Q93TW9		mxsC (3)		
	Q93TX0		mxsB (1)		
	Q8RU78		stgA (1)		
	Q8RU75		stgB (1)		
	Q8RU74		stgC (1)		
	Q8RU73		stgD (1)		
	Q8RU72		stgE (1)		
	Q8RU71		stgF (1)		
	Q8RU70		stgG (1)		
Q8RU69	stgH (1)				
Q8RU68	stgJ (1)				
<b>Cyanobacteria</b>					
<b>Chroococcales</b>					
<i>Microcystis aeruginosa</i>	Q9FD18	microcystin	mcyG (1)	preceded by NRPS	Tillett et al. 2000
	Q9FDU1		mcyD (2)		
	Q9FDU0		mcyE (1)		
<b>Nostocales</b>					
<i>Nostoc sp.</i>	Q9RAH3	nostopeptolide	nosB (1)	preceded by NRPS	Hoffmann et al. 2003

**Fig. 2.** Compilation of information concerning the published PKSII sequences used in this study. The PKSII protein name and their respective number of modules are indicated. As an example, the amphotericin synthase encompasses 6 PKSII proteins and 19 modules. Specific features in the Remarks column are detailed in the text.

## actinomycetales

## blend order



**Fig. 3.** Phylogenetic analysis of ketosynthase (KS) domains. Reconstruction was computed for 204 protein sequences on 413 sites using both the distance method (NJ;PAM matrix) and the maximum likelihood method (BIONJ; JTT matrix) with 500 bootstrap replicates. Both distance and maximum likelihood (ML) methods provided the same tree topology, suggesting that observed grouping did not result from computational artefacts. Bootstrap values are indicated as ML/NJ. Only bootstrap values higher than

70 are indicated, except when necessary. When KS domains of one polyketide synthase are grouped, they are represented as a triangle labeled with the synthesized polyketide. The module names (e.g., A, B, C, etc.) are explicitly written for the amphotericin, nystatin, or pimaricin synthases. The number of amino acid substitutions is proportional to the length of the scale. An asterisk marks the actinomycetales node. The entire tree is available at our Web site (<http://web.libragen.com/Phylogeny>).

### Exceptions to KS Clustering from One Actinomycetale PKSI

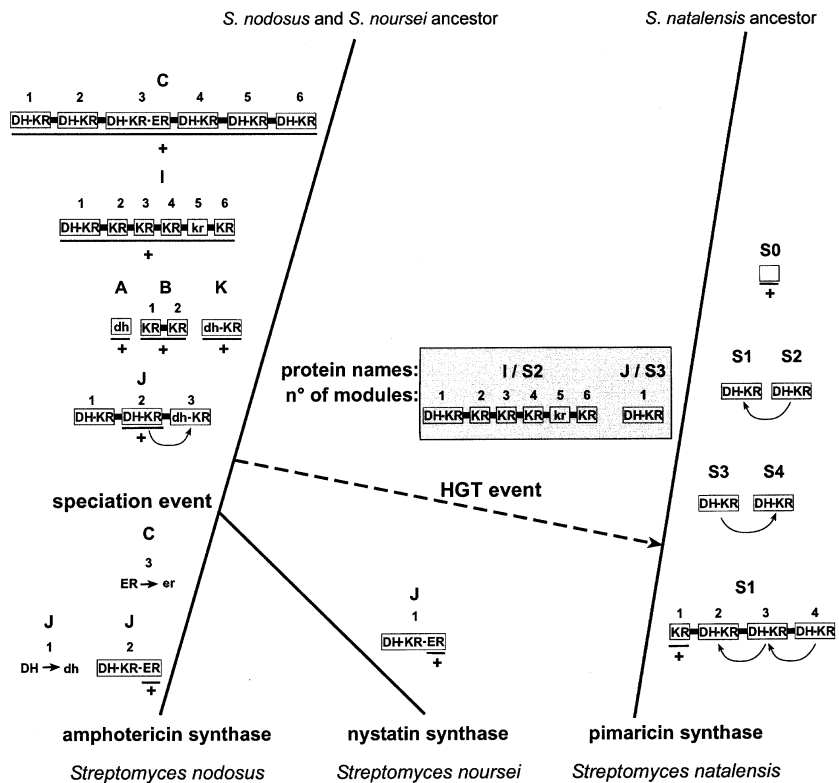
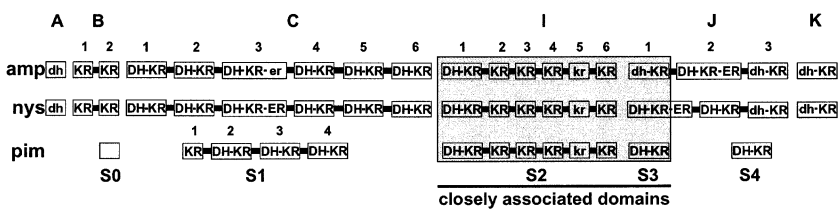
The majority of analyzed KS domains are clustered under a reliable node (Fig. 3; marked with an asterisk; 100/97 as ML/NJ bootstrap values). All these KS domains appear to be encoded by actinomycetale species. The rest of the KS domains are located within a blend group including *Nostocales*, *Chroococcales*, *Myxococcales*, and *Pseudomonadales* species (Fig. 2). In most cases, the KS domains of a given actinomycetale PKSI were clustered as a monophyletic group. As an example, 12 KS domains from the avermectin synthase (aveA1 to aveA4; Fig. 2) are clustered. Such groupings of KS domains are only supported within the actinomycetales group, as previously reported for a shorter set of 92 KS domains (Lopez 2003). Several exceptions to the rule of KS domain clustering in actinomycetales were observed, however. The KS domains of ascomycin (fkb), rapamycin (rap), and rifamycin (rif) synthases showed an imperfect clustering pattern since the KS domains from fkbA4, rapC4, and rifA2 did not fit these groupings (Figs. 3 and 5A). Moreover, in the case of mycinamicin synthase, four KS domains, mycA12, mycA13, mycA14, and mycA15, are clustered apart from the other mycA KS domains (Fig. 3). The four KS gene regions exhibit a low GC content (66%) for a *Streptomyces* gene (Fig. 2). Two other examples of phylogenetic incongruencies concern pikA13 (pikro-

mycin) and rifA2 (rifamycin) (Fig. 3), which fall outside their PKSI group.

### The Particular Case of Amphotericin, Nystatin, and Pimaricin Synthases

The amphotericin (amp) and nystatin (nys) synthases are very similar (Caffrey et al. 2001). Their modular organizations are identical (Fig. 4) and most of the domains are close pairwise (Fig. 5). In other words, considering the KS, AT, and ACP phylogenies, nysA is strongly clustered with ampA and not with other nys domains (Fig. 5). The additional domains of the reductive loop are mostly identical, even when the domains are nonfunctional (Fig. 4). The second to the sixth modules of nysI PKS were originally described with additional dehydratase domains (DH), which were noted to be nonfunctional due to large deletions (Brautaset et al. 2000). However, these domains were probably erroneously annotated as DH domains instead of AT-KR linkers (Yadav et al. 2003). Thus, the reductive loops (i.e., additional domains DH, KR, and ER; Fig. 1) of amp and nys synthases are identical. The amphotericin and nystatin synthases are mostly symmetrical (Fig. 4; top).

Among the five PKSI of the pimaricin synthase, only the KS domains from pimS2 and pimS3 appear clustered with the amp/nys KS domains. All the KS domains from ampI, nysI, pimS2 and KS domains from ampJ2, nysJ2, pimS3 are clustered as trios



**Fig. 4.** Hypothetical evolution for the amphotericin, nystatin, and pimicic acid synthases. The three polyketide synthases are aligned at the top and their putative evolutionary history is proposed below. A rectangle represents a PKS module with the implied necessary KS, AT, and ACP domains, which are not labeled. In other words, the empty rectangle of pimS0 represents one module containing the KS–AT–ACP domains. The additional reductive domains are indicated inside each rectangle as dehydratase (DH), ketoreductase (KR), or enoylreductase (ER). The inactive domains are labeled within lowercase letters. The gray rectangle highlights the domains that are closely associated with the KS, AT, and ACP phylogenies (Fig. 5). The curved arrows indicate hypothetical duplications based on phylogenies, and the plus signs, the addition of units.

(Fig. 5A). These modules are closely associated (Fig. 4) within the KS, AT, and ACP domain phylogenies with high bootstrap values (Fig. 5).

#### *The Soraphen Synthase is Separated Between Two Bacterial Orders*

Three KS domains from the soraphen (sor) synthase of a myxococcale (*Sorangium cellulosum*) were clustered in the actinomycetales group, whereas the four remaining domains were strongly clustered in the blend group (Fig. 3).

#### *KS with Unusual Functions*

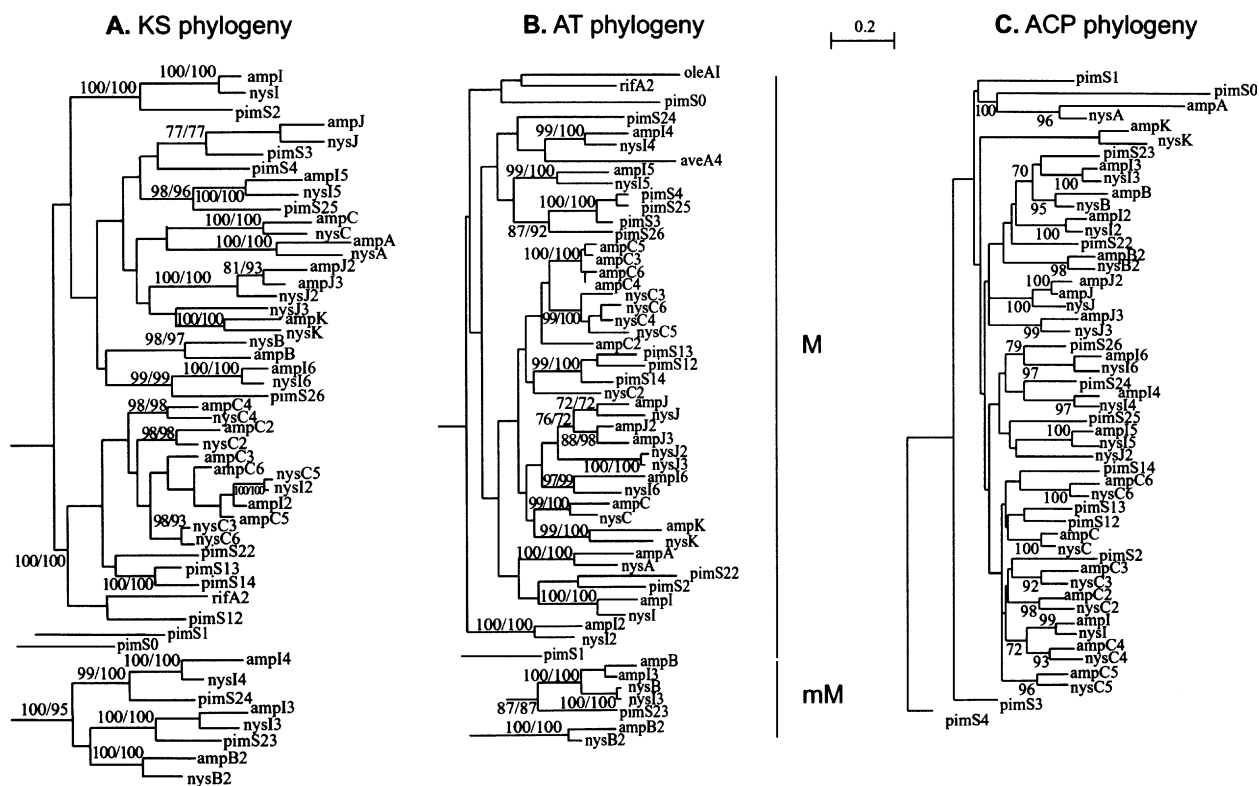
The active site of KS domains is usually a cysteine. This amino acid can be replaced by a glutamine or a serine, leading to the KS<sup>Q</sup> or KS<sup>S</sup> domain, respectively. The KS<sup>Q</sup> domains lack condensation activity and were observed grouped apart from their PKS (Fig. 3). The KS<sup>S</sup> domains also lack condensation activity and their decarboxylase activity is low

(Caffrey et al. 2001). They were observed inside their polyketide synthase of origin (Fig. 5A).

The KS associated with 2 AT merged in one loading module have to condense two substrates instead of one (Ligon et al. 2002) and were grouped with KS<sup>Q</sup>. Lastly, the KS domains involved in hybrid NRPS/PKSI structures condense an acyl group onto an amino acid chain and were all clustered, whatever their origin (Fig. 3).

#### **Discussion**

According to the hypothesis of PKS gene evolution, modular organization and chemical diversity of polyketides produced by PKS should have arisen from duplication of domains (Hopwood 1997; Zhu et al. 2002). Thus, more sequence similarity is expected between the domains of a given PKS than those of different ones (Gokhale and Khosla 2000) if the time elapsed between duplication and speciation is not too long. In this last case, the pattern observed would certainly be two closely associated clusters



**Fig. 5.** Phylogenetic subtrees for the amphotericin, nystatin, and pimarinic PKS synthases based on (A) KS domains, (B) AT domains, and (C) ACP domains. The two subtrees from KS and AT phylogenies are extracted from the complete reconstruction described in Materials and Methods. The ACP phylogeny was computed (NJ, PAM) with the 51 domains from the three polyketide synthases on 84 sites. The substrate incorporation for AT domains is indicated as malonyl (M) and methylmalonyl (mM).

such as the amphotericin and nystatin synthases. Indeed, many PKSIs computed in this study support this expectation (Figs. 3 and 5). This indicates that PKSIs in actinomycetales are more homogeneous than in other orders. This assumption was confirmed by comparison of the two means of pairwise differences between the actinomycetales group ( $n = 12,403$ , mean = 0.4788) and the blend group ( $n = 1081$ , mean = 0.7979), which were significantly different by Mann–Whitney test ( $p < 2.2 \times 10^{-16}$ ). This homogeneity may result from a high frequency of HGT within the actinomycetales group.

Several possible examples of HGT events were observed, especially between actinomycetales species. HGT would ensure genetic coherence and slow diversification between related bacteria (Torsvik et al. 2002). In the case of closely related species, frequent HGT events might also create similarities as vertical inheritance (Gogarten et al. 2002). Hence, HGT is a likely explanation for the homogeneity of the actinomycetales group.

#### *Phylogenetic Incongruencies Reveal HGT Events*

Evidence for the transfer of single domains was observed as for the KS domains *fkbB4* and *rapC4*, which are separated from their PKSIs of origin

(Fig. 3). Rapamycin and ascomycin synthases are PKSIs produced by two strains of *Streptomyces hygroscopicus* and involved in the synthesis of two closely related immunosuppressive macrolides (Wu et al. 2000). Moreover, since *fkbA4* and *rapC4* are very similar, the rapamycin and ascomycin synthases could have acquired this subunit by a recent HGT (Hopwood 1997).

It was suggested that the domains from the mycinamicin synthase, whose genes display a low GC content, resulted from a HGT (Anzai et al. 2003). Since atypical sequences from HGT tend to homogenize with the receptor genetic features (Gogarten et al. 2002; Lawrence and Roth 1996), this transfer must be recent. Indeed, the GC content (66%) is not actually optimized for an actinomycetales like *Micromonospora griseorubida* (72%). Nevertheless, this GC deviation was found in the third position of codons (Anzai et al. 2003). Since the phylogenetic reconstructions were based on protein sequences, most of the deviation was blurred due to the genetic code redundancy. Therefore, the taxonomic order of the donor organism cannot be determined.

The KS domains of *Streptomyces* species were already described as a monophyletic group (Moffitt and Neilan 2003). However, the strongly supported node for the actinomycetales group also contains the KS

domains from a myxococcale, *Sorangium cellulosum* (soraphen producer) (Fig. 3). The monophyletic phylogenetic hypothesis, where all soraphen KS domains would be clustered outside the actinomycetales node (100/97), was tested with the KHT test and rejected ( $\ln L/SD = 6.29 > 1.96$ ), Schupp et al. (1995) suggested that a genetic exchange occurred between actinomycetales and *S. cellulosum*, based on their GC content and codon usage. Such a clustering pattern may support the hypothesis that half of the soraphen cluster was acquired by HGT from an actinomycetales species.

#### *HGT Versus the Alternative Evolution Hypothesis*

The main alternative hypothesis to explain the homogeneity of actinomycetales could rely on a recent origin of PKSIs for the actinomycetales, despite the fact that they exhibit great polyketide diversity. Such a hypothesis is ruled out by the analysis of the blend group. The PKSIs appear to be mixed between different bacterial phyla (Fig. 3). The common bacterial ancestor therefore probably carried a large diversity of PKSIs which predated the origin of actinomycetales. Explaining the relative homogeneity within actinomycetales by common ancestry would imply that the initial diversity was lost, then re-created through *de novo* duplications and specialization — little parsimonious hypothesis. Alternatively, we propose that numerous HGTs between actinomycetales may have secondarily homogenized PKSIs genes in this phylum. Eventually, this pattern may be due to a higher evolution rate of the PKSIs genes from nonactinomycetales species. However, no significant difference was observed between the relative evolution rates of actinomycetales and nonactinomycetales, thus excluding this hypothesis ( $p = 0.673$ ).

Another hypothesis would rely on a phylogenetic reconstruction artifact: long branch attraction. First, we computed the reconstruction by two different methods—distance and maximum likelihood—and the same topology was obtained. Second, we computed a new distance-based reconstruction without the longest branches of the hybrid group, the KS from loading modules, and the closest KS from the root: the three KSs from the soraphen synthase. The tree topology remains unchanged and the actinomycetales group is always supported by a bootstrap value of 99.

One can argue that transferred genes are stable only under positive selection. Thus, alien genes must be functional and provide a novel adaptive advantage; otherwise they are lost by random mutation (Kurland et al. 2003). Secondary metabolites like polyketides respond to that constraint: they are synthesized under selective pressures and do not modify primary metabolism. The functionality of PKSIs in a

heterologous environment was demonstrated with the production of a functional megalomicin PKSIs in *Streptomyces lividans* and *Saccharopolyspora erythraea*, despite the fact that this PKSIs gene cluster was originally described in *Micromonospora megalomicea* (Volchegursky et al. 2000; Shah et al. 2000).

#### *Putative Evolutionary History of Amphotericin, Nystatin, and Pimaricin Synthases*

The three KS domains from ampA, nysA, and pimS0 are KS<sup>S</sup>. However, even if pimS0 and nysA share a KS<sup>S</sup>, their similarities do not seem sufficient to speculate a common origin (Brautaset et al. 2000). This active site mutation could have been acquired twice independently by convergent evolution.

The close pairwise-associated domains (Fig. 4; gray rectangle) observed with the three phylogenies (Fig. 5) suggest a common origin for these three synthases. The amphotericin and nystatin synthases are highly similar, implying a recent common origin. The pimS2 PKS appears to have been integrated through a HGT from an amp or a nys gene. The *pimS2* gene shows a peculiar codon usage and third-codon position GC content compared to other pimaricin synthase genes (Aparicio et al. 2000). This region, which includes seven modules, represents a large HGT event of 34 kb.

#### *Hybrid NRPS/PKSI and Unusual KS Function*

The KS domains positioned after a NRPS module possess an unusual function that mediates the condensation of an acyl extender unit onto the amino acid chain from the previous NRPS module. This functional constraint might explain why each KS domain preceded by a NRPS module clustered within the hybrid group and not within its own PKSIs.

It has been suggested that epoB was acquired by HGT, based on several rigorous criteria (Lopez 2003). One of them is the phylogenetic incongruence, where epoB is excluded from the epothilone PKS but is clustered within the hybrid NRPS/PKS group. Lopez suggested that this hybrid cluster does not result from a common ancestor but from a convergent evolution, since several hybrids, such as KS domains from the rapamycin and ascomycin synthases, do not cluster in this group. However, these two cases revealed by Lopez (2003) are not exceptions since the hybrid group concerns only the KS preceded by a NRPS module and not positioned after PKS domains as in the rapamycin or ascomycin synthase. Nevertheless, the GC deviation, the presence of flanking transposon sequences, and the great genetic distance consistently support the hypothesis that HGT is responsible for the

occurrence of *epoB* (Lopez 2003). Moreover, the hypothesis of convergent evolution is likely to be true, since the KS domains from hybrid systems are clustered due to a functional constraint. The function of KS<sup>S</sup> domains seems to be very low processive decarboxylation activity and remains to be explained. These domains do not seem to have evolved under strong functional pressure, unlike the other unusual KS domains.

#### *Polyketide Synthase Genes May Evolve with Mobile Elements*

Since large HGTs of PKS I clustered genes have occurred between actinomycetales and not within the other bacterial orders, a biological mechanism must exist, specific to actinomycetales. We suggest that the linearity and the instability of actinomycetales chromosomes could be implicated in this mechanism.

The chromosome of *Streptomyces* species is linear and its extremities are genetically unstable since large deletions and circularization occur at significant frequencies (up to 1%) (Chen et al. 2002; Weaver et al. 2004). Many of the biosynthetic pathways of *Streptomyces coelicolor* and *Streptomyces avermitilis* were found on chromosome arms, whereas the essential genes were located in the core chromosome (Weber et al. 2003; Challis and Hopwood 2003). Linear plasmids can acquire chromosomal end genes (Wiener et al. 1998; Weaver et al. 2004) and reciprocally (Bentley et al. 2004; Weaver et al. 2004). Therefore, these gene clusters might be transferred through extrachromosomal elements within the genus *Streptomyces* and other species with a high GC content, such as actinomycetales.

The antibiotic gene clusters and their associated resistance genes are likely to be cotransferred, such as in the case of the streptomycin synthase (Egan et al. 2001; Challis and Hopwood 2003) and the tylosin gene cluster (Stonesifer et al. 1986). The streptomycin and its resistance gene were demonstrated to be transferred at a high frequency between *Streptomyces* species (Wiener et al. 1998; Egan et al. 2001). Entire biosynthetic pathways may have evolved through HGT in a similar way as did pathogenicity islands (Thompson et al. 2002).

The presence of transposon-like sequences near biosynthetic gene clusters may support genetic exchanges in the cell. Direct repeated sequences are present at both extremities of the PKS II mithramycin gene cluster (Lombó et al. 1999) and suggest a transfer through the integration of extrachromosomal DNA. Then a gene cluster similar to the mithramycin cluster was found on the linear plasmid pSLA2-L (Mochizuki et al. 2003). This hypothesis of transfer between genetic structures was also confirmed for the methylenomycin gene cluster, which can be located on a

circular plasmid, on a linear plasmid, and integrated into the *S. coelicolor* chromosome (Wiener et al. 1998; Yamasaki et al. 2003). The gene transfer of the methylenomycin gene cluster from the circular to the linear plasmid was confirmed (Yamasaki et al. 2003). Concerning PKS I genes, the tylosin gene cluster was demonstrated to be highly transferred through conjugation between *Streptomyces fradie* strains (Stonesifer et al. 1986). Under selective pressures, some *S. fradie* strains seem to have acquired the tylosin gene cluster from other strains. A *Mycobacterium ulcerans* PKS I gene cluster, which likely encodes a mycolactone PKS I, is located between two repeated insertion sequences (Jenkin et al. 2003; Stinear et al. 2004). Moreover, subtractive hybridization experiments performed between *Mycobacterium marinum* and *M. ulcerans* demonstrated that the most important DNA acquisition by *M. ulcerans* was this PKS I cluster and the two insertion sequences (Jenkin et al. 2003). This gene cluster is located on a giant plasmid, suggesting evolution through a recent horizontal acquisition followed by duplication (Stinear et al. 2004). Such an association of insertion sequences with PKS I genes is a strong argument for HGT (Lopez 2003). This first insight of PKS I associated with insertion sequences highlights the possible genetic exchanges of these biosynthetic clusters, adding further credibility to the hypothesis of frequent HGT in actinomycetales PKS I.

#### **Conclusion**

Several observations based on KS domain phylogeny suggest that frequent HGTs of PKS I genes have occurred between actinomycetales species. The homogeneity and the robustness of the actinomycetales group may result in part from these genetic transfers. The linearity and instability of actinomycetales chromosome and the many possibilities for natural conjugation and for the transfer of genetic elements could explain the high frequency of HGT. The suggestion is advanced here that their homogeneity may rely on the genetic specificity of actinomycetales.

The Selfish Operon Theory (Lawrence and Roth 1996) proposes that genes evolved clustered in operon through HGT to maintain their frequencies in the gene pool. In this theory, a selective advantage for the host is not required since genes from operons are clustered for themselves to avoid genetic drift. This model is well adapted to the PKS I gene clusters (Lopez 2003). HGT maintains the capability for numerous genetic responses to environmental stress in bacterial populations without strong environmental pressures. Therefore, numerous genetic exchanges would enable rapid adaptation if the environmental pressures appear again.



**Acknowledgment.** We are indebted to the Pôle Bioinformatique Lyonnais for the kind access to their databanks and informatic resources. We gratefully acknowledge Jean-Luc Pernodet and Ludovic Orlando for the fruitful discussions regarding the manuscript. This contribution was supported by European project METACONTROL QLK3-CT-2002-02068.

## References

- Anzai Y, Saito N, Tanaka M, Kinoshita K, Koyama Y, Kato K (2003) Organization of the biosynthetic gene cluster for the polyketide macrolide mycinamicin in *Micromonospora griseorubida*. FEMS Microbiol Lett 218:135–341
- Aparicio JF, Fouces R, Mendes MV, Olivera N, Martin JF (2000) A complex multienzyme system encoded by five polyketide synthase genes is involved in the biosynthesis of the 26-membered polyene macrolide pimaricin in *Streptomyces natalensis*. Chem Biol 7:895–905
- Bentley SD, Brown S, Murphy LD, Harris DE, Quail MA, Parkhill J, Barrell BG, McCormick JR, Santamaria RI, Losick R, Yamasaki M, Kinashi H, Chen CW, Chandra G, Jakimowicz D, Kieser HM, Kieser T, Chater KF (2004) SCP1, a 356 023 bp linear plasmid adapted to the ecology and developmental biology of its host, *Streptomyces coelicolor* A3(2). Mol Microbiol 51:1615–1628
- Brautaset T, Sekurova ON, Sletta H, Ellingsen TE, Strom AR, Valla S, Zotchev SB (2000) Biosynthesis of the polyene antifungal antibiotic nystatin in *Streptomyces noursei* ATCC 11455: analysis of the gene cluster and deduction of the biosynthetic pathway. Chem Biol 7:395–403
- Caffrey P, Lynch S, Flood E, Finnan S, Oliynyk M (2001) Amphotericin biosynthesis in *Streptomyces nodosus*. Deductions from analysis of polyketide synthase and late genes. Chem Biol 8:713–723
- Challis GL, Hopwood DA (2003) Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species. Proc Natl Acad Sci USA 100:14555–14561
- Chen CW, Huang CH, Lee HH, Tsai HH, Kirby R (2002) Once the circle has been broken: dynamics and evolution of *Streptomyces* chromosomes. Trends Genet 18:522–529
- Cortes J, Haydock SF, Roberts GA, Bevtitt DJ, Leadlay PF (1990) An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. Nature 348:176–178
- Cundliffe E, Bate N, Butler A, Fish S, Gandeche A, Merson-Davies L (2001) The tylosin-biosynthetic genes of *Streptomyces fradiae*. Antonie Van Leeuwenhoek 79:229–234
- Du L, Sanchez C, Shen B (2001) Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. Metab Eng 3:78–95
- Egan S, Wiener P, Kallifidas D, Wellington EM (2001) Phylogeny of *Streptomyces* species and evidence for horizontal transfer of entire and partial antibiotic gene clusters. Antonie Van Leeuwenhoek 79:127–133
- Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle
- Gaitatzis N, Silakowski B, Kunze B, Nordsiek G, Blocker H, Hofle G, Muller R (2002) The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase. J Biol Chem 277:13082–13090
- Galtier N, Gouy M, Gautier C (1996) SEAVIEW and PHYLO\_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput Appl Biosci 12:543–548
- Gogarten JP, Doolittle WF, Lawrence JG (2002) Prokaryotic evolution in light of gene transfer. Mol Biol Evol 19:2226–2238
- Gokhale RS, Khosla C (2000) Role of linkers in communication between protein modules. Curr Opin Chem Biol 4:22–27
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704
- Hoffmann D, Hevel JM, Moore RE, Moore BS (2003) Sequence analysis and biochemical characterization of the nostopeptolide A biosynthetic gene cluster from *Nostoc* sp. GSV224. Gene 311:171–180
- Hopwood DA (1997) Genetic contributions to understanding polyketide synthases. Chem Rev 97:2465–2497
- Ikeda H, Nonomiya T, Usami M, Ohta T, Omura S (1999) Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*. Proc Natl Acad Sci USA 96:9509–9514
- Jenkin GA, Stinear TP, Johnson PDR, Davies JK (2003) Subtractive hybridization reveals a type I polyketide synthase locus specific to *Mycobacterium ulcerans*. J Bacteriol 185:6870–6882
- Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. CABIOS 8:275–282
- Kakavas SJ, Katz L, Stassi D (1997) Identification and characterization of the niddamycin polyketide synthase genes from *Streptomyces caelestis*. J Bacteriol 179:7515–7522
- Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. J Mol Evol 29:170–179
- Kurland CG, Canback B, Berg OG (2003) Horizontal gene transfer: a critical view. Proc Natl Acad Sci USA 100:9658–9662
- Lawrence JG, Roth JR (1996) Selfish operon theory: Horizontal gene transfer may drive the evolution of gene clusters. Genetics 143:1843–1860
- Ligon J, Hill S, Beck J, Zirkle R, Molnar I, Zawodny J, Money S, Schupp T (2002) Characterization of the biosynthetic gene cluster for the antifungal polyketide soraphen A from *Sorangium cellulosum* So ce26. Gene 285:257–267
- Lombó F, Braña AF, Méndez C, Salas JA (1999) The mithramycin gene cluster of *Streptomyces argillaceus* contains a positive regulatory gene and two repeated DNA sequences that are located at both ends of the cluster. J Bacteriol 181:642–647
- Lopez JV (2003) Naturally mosaic operons for secondary metabolite biosynthesis; variability and putative horizontal transfer of discrete catalytic domains of the epothilone polyketide synthase locus. Mol Gen Genomics 270:420–431
- Mochizuki S, Hiratsu K, Suwa M, Ishii T, Sugino F, Yamada K, Kinashi H (2003) The large linear plasmid pSLA2-L of *Streptomyces rochei* has an unusually condensed gene organization for secondary metabolism. Mol Microbiol 48:1501–1510
- Moffitt MC, Neilan BA (2003) Evolutionary affiliations within the superfamily of ketosynthases reflect complex pathway associations. J Mol Evol 56:446–457
- Molnar I, Schupp T, Ono M, Zirkle R, Milnamow M, Nowak-Thompson B, Engel N, Toupet C, Stratmann A, Cyr DD, Gorchach J, Mayo JM, Hu A, Goff S, Schmid J, Ligon JM (2000) The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* So ce90. Chem Biol 7:97–109
- Nowak-Thompson B, Gould SJ, Loper JE (1997) Identification and sequence analysis of the genes encoding a polyketide synthase required for pyoluteorin biosynthesis in *Pseudomonas fluorescens* Pf-5. Gene 204:17–24
- Perrière G, Gouy M (1996) WWW-Query: an on-line retrieval system for biological sequence banks. Biochimie 78:364–369

- Robinson-Rechavi M, Huchon D (2000) RRTree: Relative-rate tests between groups of sequences on a phylogenetic tree. *Bioinformatics* 16:296–297
- Schupp T, Toupet C, Cluzel B, Neff S, Hill S, Beck JJ, Ligon JM (1995) A *Sorangium cellulosum* (myxobacterium) gene cluster for the biosynthesis of the macrolide antibiotic soraphen A: cloning, characterization, and homology to polyketide synthase genes from actinomycetes. *J Bacteriol* 177:3673–3679
- Schwecke T, Aparicio JF, Molnar I, Konig A, Khaw LE, Haydock SF, Oliynyk M, Caffrey P, Cortes J, Lester JB, Bohm GA, Stauton J, Leadlay PF (1995) The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc Natl Acad Sci USA* 92:7839–7843
- Shah S, Xue Q, Tang L, Carney JR, Betlach M, McDaniel R (2000) Cloning, characterization and heterologous expression of a polyketide synthase and P-450 oxidase involved in the biosynthesis of the antibiotic oleandomycin. *J Antibiot* 53:502–508
- Shen B (2003) Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr Opin Chem Biol* 7:285–295
- Silakowski B, Schairer HU, Ehret H, Kunze B, Weinig S, Nordsiek G, Brandt P, Blocker H, Hofle G, Beyer S, Muller R (1999) New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1. *J Biol Chem* 274:37391–37399
- Silakowski B, Nordsiek G, Kunze B, Blocker H, Muller R (2001) Novel features in a combined polyketide synthase/non-ribosomal peptide synthetase; the myxalamid biosynthetic gene cluster of the myxobacterium *Stigmatella aurantiaca* Sga15. *Chem Biol* 8:59–69
- Staunton J, Weissman KJ (2001) Polyketide biosynthesis: a millennium review. *Nat Prod Rep* 18:380–416
- Stinear TP, Mve-Obiang A, Small PLC, Frigui W, Pryor MJ, Brosch R, Jenkin GA, Johnson PDR, Davies JK, Lee RE, Adusumilli S, Gamier T, Haydock SF, Leadlay PF, Cole ST (2004) Giant plasmid-encoded polyketide synthases produce the macrolide toxin of *Mycobacterium ulcerans*. *Proc Natl Acad Sci USA* 101:1345–1349
- Stonesifer J, Matsushima P, Baltz RH (1986) High frequency conjugal transfer of tylosin genes and amplifiable DNA in *Streptomyces fradiae*. *Mol Gen Genet* 202:348–355
- Tang L, Yoon YJ, Choi CY, Hutchinson CR (1998) Characterization of the enzymatic domains in the modular polyketide synthase involved in rifamycin B biosynthesis by *Amycolatopsis mediterranei*. *Gene* 216:255–265
- Thompson CJ, Fink D, Nguyen LD (2002) Principles of microbial alchemy: insights from the *Streptomyces coelicolor* genome sequence. *Genome Biol* 3:1–4
- Thompson JD, Plewniak F, Thierry J-C, Poch O (2000) DBClustal: rapid and reliable global multiple alignments of protein sequences detected by database searches. *Nucleic Acids Res* 28:2919–2926
- Tillett D, Dittmann E, Erhard M, von Dohren H, Borner T, Neilan BA (2000) Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide-polyketide synthetase system. *Chem Biol* 7:753–764
- Torsvik V, Ovreas L, Thingstad TF (2002) Prokaryotic Diversity—Magnitude, dynamics, and controlling factors. *Science* 296:1064–1066
- Volchegursky Y, Hu Z, Katz L, McDaniel R (2000) Biosynthesis of the anti-parasitic agent megalomicin: transformation of erythromycin to megalomicin in *Saccharopolyspora erythraea*. *Mol Microbiol* 37:752–762
- Waldron C, Matsushima P, Rosteck PR Jr, Broughton MC, Turner J, Madduri K, Crawford KP, Merlo DJ, Baltz RH (2001) Cloning and analysis of the spinosad biosynthetic gene cluster of *Saccharopolyspora spinosa*. *Chem Biol* 8:487–499
- Watanabe K, Wang CCC, Boddy CN, Cane DE, Khosla C (2003) Understanding substrate specificity of polyketide synthase modules by generating hybrid multimodular synthases. *J Biol Chem* 278:42020–42026
- Weaver D, Karoonuthaisiri N, Tsai HH, Huang CH, Ho ML, Gai S, Patel KG, Huang J, Cohen SN, Hopwood DA, Chen CW, Kao CM (2004) Genome plasticity in *Streptomyces*: identification of 1 Mb TIRs in the *S. coelicolor* A3(2) chromosome. *Mol Microbiol* 51:1535–1550
- Weber T, Welzel K, Pelzer S, Vente A, Wohlleben W (2003) Exploiting the genetic potential of polyketide producing *Streptomyces*. *J Biotechnol* 106:221–223
- Wiener P, Egan S, Wellington MH (1998) Evidence for transfer of antibiotic-resistance genes in soil populations of *Streptomyces*. *Mol Ecol* 7:1205–1216
- Wu K, Chung L, Revill P, Katz L, Reeves CD (2000) The FK520 gene cluster of *Streptomyces hygroscopicus* var. *ascomyceticus* (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. *Gene* 251:81–90
- Xue Y, Wilson D, Sherman DH (2000) Genetic architecture of the polyketide synthases for methymycin and pikromycin series macrolides. *Gene* 245:203–211
- Yadav G, Gokhale RS, Mohanty D (2003) Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *J Mol Biol* 328:335–363
- Yamasaki M, Ikuto Y, Ohira A, Chater K, Kinashi H (2003) Limited regions of homology between linear and circular plasmids encoding methylenomycin biosynthesis in two independently isolated streptomycetes. *Microbiology* 149:1351–1356
- Zhu G, LaGier MJ, Stejskal F, Millership JJ, Cai X, Keithly JS (2002) *Cryptosporidium parvum*: the first protist known to encode a putative polyketide synthase. *Gene* 298:79–89