J Ind Microbiol Biotechnol (2012) 39:1541–1547
DOI 10.1007/s10295-012-1149-2

Horizontal gene transfer and gene conversion drive evolution of modular polyketide synthases

Jurica Zucko · Paul F. Long · Daslav Hranueli · John Cullum

Received: 5 February 2012 / Accepted: 8 May 2012 / Published online: 30 May 2012

Abstract Soil bacteria live in a very competitive environment and produce many secondary metabolites; there appears to be strong selective pressure for evolution of new compounds. Secondary metabolites are the most important source of chemical structures for the pharmaceutical industry and an understanding of the evolutionary process should help in finding novel chemical entities. Modular polyketide synthases are a particularly interesting case for evolutionary studies, because much of the chemical structure can be predicted from DNA sequence. Previous evolutionary studies have concentrated on individual modules or domains and were not able to study the evolution of orthologues. This study overcame this problem by considering complete clusters as "organisms", so that orthologous modules and domains could be identified and used to characterise evolutionary pathways. Seventeen modular polyketide synthase clusters were identified that fell into six classes. Gene conversion within clusters was very common (affecting about 15 % of domains) and was detected by discordance in phylogenetic trees. An evolutionary model is proposed in which a single cross over between two different clusters (i.e. horizontal gene transfer) would generate a cluster of very different architecture with radically different chemical products; subsequent gene conversion and deletions would explore chemical variants. Two probable examples of such recombination were found. This model suggests strategies for detecting horizontal gene transfer in cluster evolution.

Keywords Evolution · Horizontal gene transfer · Gene conversion · Polyketide synthases

Introduction

Secondary metabolites are the most important source of pharmaceuticals. They are produced by many organisms and are especially prevalent in soil microorganisms. A striking observation is that the biosynthesis genes for a particular metabolite occur together in a cluster. Such clusters of coevolving genes are fascinating material for evolutionary studies [2, 5]. The large number of structurally diverse secondary metabolites observed in nature suggests that there is rapid evolution of these metabolites with selection for chemical diversity. Their postulated roles, as antibiotics to give the producer competitive advantage, or as signalling molecules [4] to coordinate members of a clone, would result in frequency-dependent selection and, thus, diversity. A better understanding of the evolution of secondary metabolites will help in the search for novel chemical entities.

Modular polyketide synthases (PKSs) in actinomycetes are a particularly attractive system for evolutionary studies. The polyketide backbones are assembled in a series of synthesis steps each carried out by a separate module,

Electronic supplementary material The online version of this article (doi:10.1007/s10295-012-1149-2) contains supplementary material, which is available to authorized users.

J. Zucko · J. Cullum (✉)
Department of Genetics, University of Kaiserslautern, Postfach 3049, 67653 Kaiserslautern, Germany
e-mail: cullum@rhrk.uni-kl.de

J. Zucko · D. Hranueli
Faculty of Food Technology and Biotechnology, University of Zagreb, Pierottijeva 6, 10000 Zagreb, Croatia

P. F. Long
Department of Chemistry, Institute of Pharmaceutical Science, King's College London, Franklin–Wilkins Building, Stamford Street, London SE1 9NH, UK

which contains several domains; many biosynthesis clusters contain 10 or more modules. To a large extent, the chemistry of the product can be predicted from the DNA sequence [21] by analysing the domains present. Because domains of a particular type share a degree of sequence conservation, homologous recombination is an obvious mechanism for generating new polyketides [7, 19]. In some cases, horizontal gene transfer of secondary metabolite clusters has been deduced [3, 11] and conjugation systems are common in actinomycetes [6] so that, potentially, any pair of modular PKS clusters should have the opportunity to recombine. Modelling of recombination between clusters in silico has shown a high potential for generating novel molecules [19].

Evolutionary studies have been carried out on modules and individual domains of modular PKS clusters [4, 8, 9, 17]. The acyl transferase (AT) and the ketosynthase (KS) domains have been analysed in most detail, because they are both present in all extender modules. AT domains select the substrate and the substrate specificity can usually be predicted from the sequence. When phylogenetic trees are constructed, AT domains mostly occur in clades corresponding to their substrate specificity. KS domains couple the new substrate to the existing growing chain and do not have any known specificity. In phylogenetic trees, it is usually found that the KS domains from a particular biosynthetic cluster are more closely related to each other than to those from other clusters. As most clusters contain AT domains of differing specificity, there is discordance between the phylogenetic trees of modules based on the KS and AT domains, which suggests horizontal gene transfer [16]. These observations lead to the interesting hypothesis that modular PKS clusters evolved by duplication of an ancestral module, followed by recombination events that replaced the AT (and other domains) to generate different module specificities [8, 17]. However, an alternative contributor to the discordance in the phylogenetic trees could be gene conversion between modules of a cluster. The KS domains show higher sequence conservation than the other domains [19] so that it is plausible that gene conversion would occur here preferentially.

In addition to the AT and KS domains, each module contains an acyl carrier protein (ACP) domain, which binds the growing polyketide chain. There may be one or more reduction domains: ketoreductase (KR), dehydratase (DH) or enoyl reductase (ER). In the absence of reduction domains, a keto group is incorporated into the polyketide backbone. KR alone produces an alcohol, successive operation of KR and DH produce a C=C double bond, and the three reduction domains acting successively produce a fully reduced extension unit [4]. An active KR domain reduces the keto group on the β-carbon atom and generates a chiral site, whose stereochemistry can often be deduced from the sequence of the domain [1, 13]. However, it also controls the stereochemistry of the α-carbon atom [12, 20]. Thus, an active KR domain can have one of four different specificities corresponding to the stereochemistry at the two carbon atoms. There are also KR domains which are "inactive", i.e. the keto group is not reduced, but still control the stereochemistry of the α-carbon atom. Thus, there are six different possible specificities and these can usually be recognised from the DNA sequence.

In order to study the evolution of gene families it is important to identify orthologues and the study of the evolution of modular PKS clusters has been limited by the inability to distinguish orthologues and paralogues. In general, it is not possible to detect orthologues of PKS modules. However, there are some PKSs which have almost identical genetic and protein architectures (e.g. the spiramycin and tylosin clusters, Fig. 1). In such cases, it is clear that the clusters are homologous, i.e. have a common ancestor with similar cluster architecture. The corresponding modules and domains in the clusters are orthologues: i.e. we consider the cluster as the "organism" when defining orthologues. In this paper, we use such homologous gene clusters to examine the roles of gene conversion and horizontal gene transfer in the evolution of chemical diversity in modular PKSs.

Materials and methods

DNA and protein sequences

The DNA and protein sequences were obtained from the CSDB database (http://csdb.bioserv.pbf.hr/csdb/). The 17 PKS gene clusters used (Table 1; supplementary Table S1, Online Resource 1) contained 201 modules with 884 domains (195 KS, 200 AT, 180 KR, 109 DH and 200 ACP domains). The clusters in the database have been re-annotated with the ClustScan suite of programs [21] ensuring a uniform annotation and definition of module and domain boundaries.

Phylogenetic analysis

Phylogenetic analysis was performed on the DNA and protein sequences of domains using the MEGA 4 software package [24] for distance and maximal parsimony methods and the TREE-PUZZLE program [18] for a maximum likelihood method. The alignments of the DNA and protein sequences used are shown in Online Resource 2. DNA distances were estimated with the Tamura–Nei model [25]. Protein substitution models were the JTT matrix for MEGA 4 [10] and the WAG matrix [26] for TREE-PUZZLE. Distance trees were derived using the minimal evolution

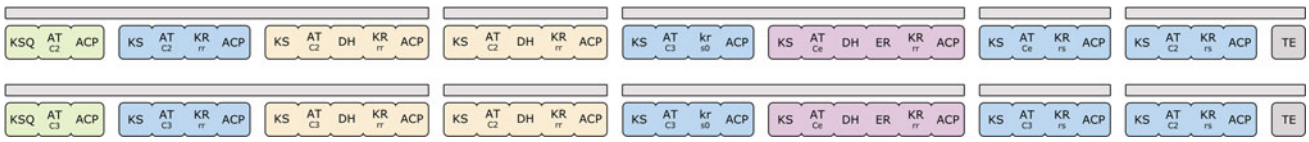

Fig. 1 The organisation of the polyketide synthase genes of the spiramycin cluster (*top*) and the tylosin cluster (*bottom*). The five genes are shown as *bars* and the eight modules (numbered *0–7*) with their constituent domains are shown as *boxes*. The specificity of domains is also shown. *AT* substrate (*C2* malonyl-CoA, *C3* methylmalonyl-CoA, *Ce* ethylmalonyl-CoA); *KR* stereochemistry of the α- and β-carbon atoms respectively (*r*, *s*, 0 if not predicted). The KSQ domain is a special domain present in some starter modules and the TE domain is responsible for release of the polyketide product from the synthase

Table 1 Seventeen PKS clusters, which have nearly identical homologues

Group	PKS gene clusters
I	Amphotericin, nystatin, pimaricin
II	Avermectin, nemadectin
III	Concanamycin, concanamycin orthologue
IV	Oligomycin, oligomycin orthologue
V	Midecamycin, niddamycin, spiramycin, tylosin
VI	Erythromycin, megalomycin, lankamycin, oleandomycin

criterion with the neighbour joining algorithm after complete deletion of gaps. One thousand bootstrap replicates were used. Maximum parsimony trees were generated using the close-neighbour-interchange method after random addition of sequences to trees [14]. The maximum likelihood analysis used quartet puzzling [22, 23] with 50,000 puzzling steps.

Results

Identifying orthologous modules

The ClustScan program [21] allows the easy uniform annotation of PKS clusters and extraction of the protein and DNA sequences of domains and modules. Re-annotation with ClustScan helped identify pairs of clusters with nearly identical organisation. The major criterion was that the clusters contained genes with the same number of modules and the corresponding modules in the two clusters had a similar structure with respect to the presence and absence of particular reduction domains. In most cases, the clusters had an identical number of genes and modules. An example of such a cluster pair is spiramycin and tylosin (Fig. 1). In the cartoon, it can be seen that the distribution of modules between genes is identical and the occurrence of reduction domains (KR, DH and ER) in modules is identical. In some cases, there were several clusters with nearly identical organisation (Table 1). For instance, the midecamycin and niddamycin clusters (group V) also have a similar structure to the spiramycin and tylosin clusters. A total of 17 clusters organised into 6 different groups could be identified.

Differences in domain function and specificity between homologous clusters

Figure 1 shows that the spiramycin and tylosin clusters differ in the specificity of several AT domains (in modules 0, 1, 2 and 6). However, the other two clusters of group V (midecamycin and niddamycin) have AT specificities identical to those of spiramycin. The changing of the AT specificity involves the replacement of several specificity-determining amino acids [21] and seems more likely to result from recombination events than mutation.

Tylosin represents a relatively unusual situation in the 17 clusters studied as four AT domains have a different specificity to those in other members of the group. The only other cases of differing AT specificity were in group IV: in modules 4 and 10, the oligomycin PKS has a C_3 specificity, whereas the oligomycin orthologue PKS has a C_2 specificity.

The KR domains were annotated for stereochemical specificity. No clear cut changes of specificity were found in the cluster groups; in a few cases, KR domains were present that could not be assigned to a stereochemical class, whereas a corresponding domain in a homologous cluster could be determined.

Gene conversion

As mentioned above, the easiest way to explain changes in AT specificity in the tylosin cluster compared to the other three clusters of group V would be the occurrence of homologous recombination with AT domains of the appropriate specificity. If a suitable domain is present in the cluster involved, then gene conversion would be an attractive mechanism. If the tylosin cluster were derived from a cluster with the same AT specificities as the other members of group V, the only donor module for the conversion of C_2 specificities to C_3 would be module 4 (Fig. 1). The DNA sequence of module 4 was compared to the sequences of those modules containing putative conversion events for AT specificity. In each case (Fig. 2) there was a long stretch (over 1 kb) of nearly identical DNA. These DNA regions contained the complete AT domain and sometimes extended into the KS domain.

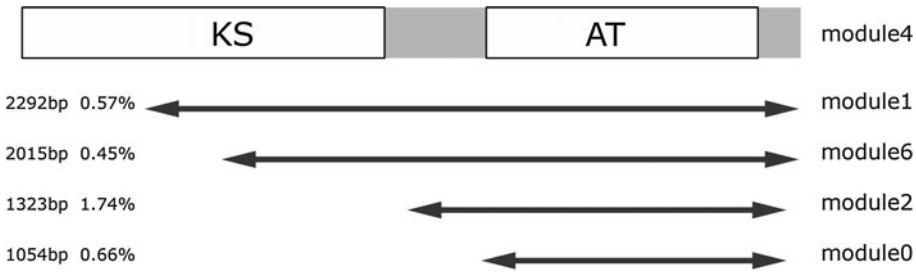

Fig. 2 Potential gene conversion events affecting AT domains in the tylosin PKS. The DNA regions of near identity with module 4 are shown for the four potential conversion events. The length of the region of near identity and the percentage of non-identical bases are also shown

The other three modules, which have other AT specificities, showed no substantial regions of identical DNA sequence with module 4. Modules 3 and 7, which both have a C_2 AT specificity, also do not show any extensive regions of DNA sequence identity.

The tylosin cluster was unusual in showing long regions of near identity between AT domains. Only two other examples were found (avermectin modules 1 and 7, spiramycin loading domain and module 1). It is to be expected that, unless the gene conversion occurred recently, subsequent sequence divergence will make it difficult to detect gene conversion using this approach. In addition, gene conversion events, which involve domains of identical specificity, are difficult to characterise by the approach used above, as the sequence divergence between the domains is smaller. A more sensitive method is to look for discordance in phylogenetic trees. The clusters in a group are assumed to be descendants of a common ancestor cluster so that the corresponding modules can be viewed as orthologues. This means that if phylogenetic trees are constructed for the domains in a particular group, the orthologues should form clades i.e. be closer in sequence to each other than to other domains in the clusters. Gene conversion, if extensive enough, may result in a module being more similar to another module in its own cluster than the orthologous modules. ClustScan [21] was used to extract the DNA and protein sequences of all domains for each group of clusters for phylogenetic analysis using three methods (a distance method, a maximum parsimony method and a maximum likelihood method). The phylogenetic trees constructed from the DNA sequences are shown in Online Resource 3. The results with all three methods were comparable. Gene conversion events were deduced from the tree topologies. Figure 3 shows an example of such an analysis for KS domains of the amphotericin and nystatin clusters. The KS domains of modules 1 and 18 form clades as expected for orthologues. However, the amphotericin KS domain from module 17 is grouped with the amphotericin KS domain from module 16 instead of the nystatin KS domain from module 17. This suggests that gene conversion has transferred sequences from amphotericin module 16 to module 17 after the two clusters diverged from each other.

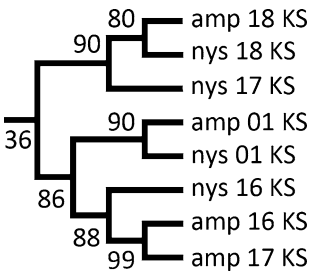

Fig. 3 Part of a phylogenetic tree for the KS domains of the amphotericin and nystatin clusters from group I. The tree was constructed using neighbour joining with a minimal evolution criterion. The bootstrap support values for the tree are shown as percentages. Complete phylogenetic trees based on DNA sequences are shown in Online Resource 3

Table 2 shows that gene conversion events in PKS clusters are not rare with more than 15 % of domains showing detectable conversion. The detailed results are shown in supplementary Tables S2–S4 (Online Resource 4). With the exception of KS domains, there was little difference between the analyses using DNA and protein sequences. In all cases, more conversion events were detected using DNA sequences than protein sequences, but apart from the KS domains, where 19 conversion events were missed, the differences were not large. The frequency of conversion events differed considerably between the different domains. It is intriguing to note that the decline in frequencies corresponds to the domain order in modules (KS–AT–KR–DH–ACP).

Table 2 Gene conversion events detected using phylogenetic trees

Domain	Total	Conversion (protein)	%	Conversion (DNA)	%
KS	195	34	17	53	27
AT	200	53	27	54	27
KR	180	32	18	35	19
DH	109	9	8	12	11
ACP	200	10	5	11	6
Total	884	136	16	165	19

Evidence for recombination events with other clusters

Most of the clusters in a group had an identical number of genes and modules. However, in group II, avermectin and nemadectin differ in having 13 and 14 modules respectively. Figure 4a shows that the nemadectin cluster possesses an additional gene carrying the loading domain. In principle, the avermectin cluster could have been derived from a cluster similar to nemadectin by fusion of the first two genes with elimination of nemadectin module 1. However, the loading domains differ and this suggests an alternative explanation that homologous recombination with another cluster had replaced the loading domain. This would imply a recombination event at one of the points indicated by the red arrows in Fig. 4a. The pimaricin cluster (Fig. 4b) in group I also differs in the loading domain region from the other two clusters in the group. In this case, the clusters cannot be related by a simple deletion event, but rather by recombination with another cluster. In addition to the differences in the loading domain region, the pimaricin cluster also has deletion of later modules in comparison to the nystatin cluster.

Discussion

The identification of orthologous domains in homologous clusters allows the deeper analysis of the evolution of PKS modules. It is striking that gene conversion is common (Table 2). In most cases, it made little difference whether DNA or protein sequences are used. However, for KS domains, considerably more conversion events were detected with DNA. This probably results from the fact that KS domains are the most highly conserved at the protein level and examination of substitution rates for synonymous and non-synonymous codons indicate that they are subjected to intensive purifying selection [28]. AT and KS domains show the highest level of conversion and the gene conversion frequencies in the other domains fall off with the order of the domains along the module (KS–AT–KR–DH–ACP). Gene conversion could explain the observation that the KS domains of a particular cluster are usually grouped together in a phylogenetic tree [8, 17]. However, given the discordance with the phylogenetic distribution of the AT domains, it was initially surprising that the AT domains showed a similar frequency of gene conversion to KS domains (Table 2). The AT domains are present in clades corresponding to the substrate specificity and most of the gene conversion events involved AT domains with the same substrate specificity. When gene conversion results in a change of substrate specificity, the whole AT domain is exchanged (Fig. 2). AT domains seem to be subject to strong purifying selection [8, 28] so that shorter gene conversion events may result in non-functional proteins. The results we obtained show the evolutionary potential for clusters without interaction with other clusters. Gene conversion can change the specificity of modules. In the case of the AT domains analysed (Figs. 1, 2) this results in replacing a hydrogen atom with a methyl group. Gene conversion or deletion events involving reduction domains can also result in chemical variety.

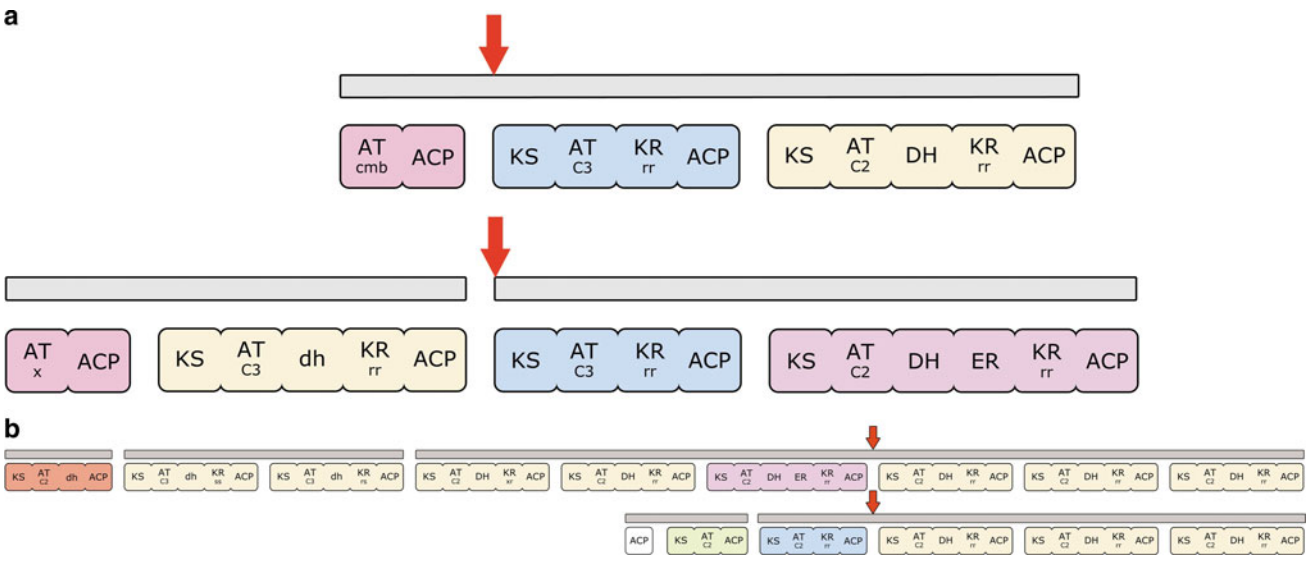

Fig. 4 Potential recombination events with other clusters. **a** The early biosynthetic modules of the avermectin (*top*) and nemadectin (*bottom*) cluster are shown. The *arrows* indicate the site where recombination with a third cluster could have occurred. **b** The early biosynthetic modules of the nystatin (*top*) and pimaricin (*bottom*) cluster. The *arrows* indicate the site where recombination with a third cluster could have occurred. Genes, modules and domains are shown as in Fig. 1

The potential chemical space that can be derived from a particular cluster is limited by the choice of domains and domain specificities that are present. Because of the strong purifying selection [28], it is unlikely that different specificities could be developed by point mutations. Another source of new specificities would be recombination with other clusters. Single cross overs would generate new cluster architectures with one end derived from each parent and modelling of homologous recombination suggested that many such recombinant clusters should give rise to polyketide products [19]. Such events should be favoured in genera such as *Streptomyces*, where the chromosomes are linear and many linear plasmids are present. Recombination by single cross overs between chromosomes and linear plasmids have been observed [15, 27] and conjugation systems are common, facilitating transfer of clusters between strains and opening up the potential for recombination between any cluster pair. *Streptomyces* are known as prolific producers of secondary metabolites, which probably arises both from their genetic potential to generate chemical diversity and the selection pressures in the soil habitat. We found two examples which probably derive from such single cross over recombination events (Fig. 4).

This suggests a new evolution model for modular biosynthetic clusters, which explains the generation of chemical diversity as well as the observation that the KS domains of a cluster group together in phylogenetic trees (Fig. 5). Single cross over recombination (i.e. horizontal gene transfer between the cluster "organisms") would generate novel cluster architectures with chemical products radically different from the parental cluster products. The chemical products would be immediately subjected to selection. Gene conversion or deletion of reduction domains would generate minor changes in the chemistry, allowing optimisation of the product. If novel cluster architectures arose from duplication of a single module [8], the extender units would all exhibit the same chemistry. If subsequent evolution by recombination with other clusters led to a diverse module chemistry, as observed in many natural clusters, it would be necessary to assume that each recombination event resulted in a product which had a selective advantage for the host. The chemical changes would often affect the degree of reduction and, thus, the ring closures that occur, so that the series of chemical products would be very different from each other. Although duplication probably does play a role in the evolution of modular PKSs [17], it seems unlikely to generate clusters with the observed combinations of module chemistry.

Single cross over recombination is likely to be a major driving force in the evolution of chemical diversity in modular biosynthetic clusters. A major challenge is how to recognise such events. The approach used in this paper to identify clusters of nearly identical structure will fail to recognise many interesting cases. After the initial recombination event, gene conversion will result in the KS domains derived from the two parent clusters becoming similar to each other. However, in many cases, the KS domains will still cluster close to those of one of the parental clusters in phylogenetic trees. Comparison of the cluster architectures may identify a series of contiguous modules that possess similar specificities in the two clusters and allow the recombination events to be reconstructed. The rapid accumulation of DNA sequencing data and the uniform annotation with the ClustScan program [21] will greatly aid this process. We propose that a module be considered the unit for evolutionary change in PKSs and that recombination between modules is the major driving force that generates chemical novelty in modular PKS clusters.

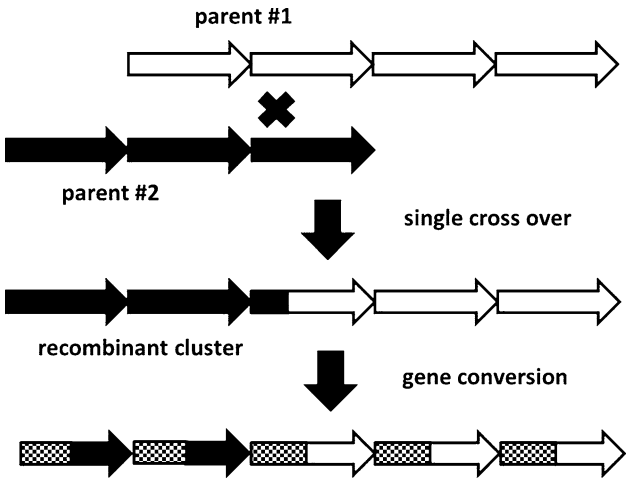

Fig. 5 Evolution model for modular PKS clusters. Two parent clusters are shown with the DNA encoding four and three modules respectively (many clusters have 10–20 modules). A single cross over generates a new recombinant cluster with five modules. As the number and specificity of the modules are different from either of the parent clusters, this can generate radically different chemical structures in the product. Subsequent gene conversion results in a homogenisation of the sequences of the KS domains so that they group together in phylogenetic trees

Acknowledgments This work was funded by a cooperation grant of the German Academic Exchange Service (DAAD) and the Ministry of Science, Education and Sports, Republic of Croatia (to J.C. and D.H.) and by the grant 058-0000000-3475 (to D.H.) from the Ministry of Science, Education and Sports, Republic of Croatia.

References

1. Castonguay R, He W, Chen AY, Khosla C, Cane DE (2007) Stereospecificity of ketoreductase domains of the 6-deoxyerythronolide B synthase. J Am Chem Soc 129:13758–13769. doi:10.1021/ja0753290

2. Demain AL (2009) Antibiotics: natural products essential to human health. Med Res Rev 29:821–842. doi:10.1002/med.20154
3. 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. doi:10.1023/A:1010296220929
4. Fischbach MA, Walsh CT (2006) Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. Chem Rev 106:3468–3496. doi:10.1021/cr0503097
5. Fischbach MA, Walsh CT, Clardy J (2008) The evolution of gene collectives: how natural selection drives chemical innovation. Proc Natl Acad Sci U S A 105:4601–4608. doi:10.1073/pnas.0709132105
6. Hopwood DA (2006) Soil to genomics: the *Streptomyces* chromosome. Annu Rev Genet 40:1–23. doi:10.1146/annurev.genet.40.110405.090639
7. Hranueli D, Cullum J, Basrak B, Goldstein P, Long PF (2005) Plasticity of the streptomyces genome-evolution and engineering of new antibiotics. Curr Med Chem 12:1697–1704. doi:10.2174/0929867054367176
8. Jenke-Kodama H, Börner T, Dittmann E (2006) Natural biocombinatorics in the polyketide synthase genes of the actinobacterium *Streptomyces avermitilis*. PLoS Comput Biol 2:e132. doi:10.1371/journal.pcbi.0020132
9. Jenke-Kodama H, Sandmann A, Müller R, Dittmann E (2005) Evolutionary implications of bacterial polyketide synthases. Mol Biol Evol 22:2027–2039. doi:10.1093/molbev/msi193
10. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282. doi:10.1093/bioinformatics/8.3.275
11. Jørgensen H, Fjaervik E, Hakvåg S, Bruheim P, Bredholt H, Klinkenberg G, Ellingsen TE, Zotchev SB (2009) Candicidin biosynthesis gene cluster is widely distributed among *Streptomyces* spp. isolated from the sediments and the neuston layer of the Trondheim fjord, Norway. Appl Environ Microbiol 75:3296–3303. doi:10.1128/AEM.02730-08
12. Keatinge-Clay AT (2007) A tylosin ketoreductase reveals how chirality is determined in polyketides. Chem Biol 14:898–908. doi:10.1016/j.chembiol.2007.07.009
13. Keatinge-Clay AT, Stroud RM (2006) The structure of a ketoreductase determines the organization of the β-carbon processing enzymes of modular polyketide synthases. Structure (Lond) 14:737–748. doi:10.1016/j.str.2006.01.009
14. Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, New York
15. Pandza S, Biuković G, Paravić A, Dadbin A, Cullum J, Hranueli D (1998) Recombination between the linear plasmid pPZG101 and the linear chromosome of *Streptomyces rimosus* can lead to exchange of ends. Mol Microbiol 28:1165–1176. doi:10.1046/j.1365-2958.1998.00877.x
16. Poptsova M (2009) Testing phylogenetic methods to identify horizontal gene transfer. Methods Mol Biol 532:227–240. doi:10.1007/978-1-60327-853-9_13
17. Ridley CP, Lee HY, Khosla C (2008) Evolution of polyketide synthases in bacteria. Proc Natl Acad Sci U S A 105:4595–4600. doi:10.1073/pnas.0710107105
18. Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18:502–504. doi:10.1093/bioinformatics/18.3.502
19. Starcevic A, Diminic J, Zucko J, Elbekali M, Schlosser T, Lisfi M, Vukelic A, Long PF, Hranueli D, Cullum J (2011) A novel docking domain interface model that can predict recombination between homoeologous modular biosynthetic gene clusters. J Ind Microbiol Biotechnol 38:1295–1304. doi:10.1007/s10295-010-0909-0
20. Starcevic A, Jaspars M, Cullum J, Hranueli D, Long PF (2007) Predicting the nature and timing of epimerisation on a modular polyketide synthase. ChemBioChem 8:28–31. doi:10.1002/cbic.200600399
21. Starcevic A, Zucko J, Simunkovic J, Long PF, Cullum J, Hranueli D (2008) ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. Nucleic Acids Res 36:6882–6892. doi:10.1093/nar/gkn685
22. Strimmer K, Goldman N, von Haeseler A (1997) Bayesian probabilities and quartet puzzling. Mol Biol Evol 14:210–213
23. Strimmer K, von Haeseler A (1996) Quartet puzzling: a quartet maximum–likelihood method for reconstructing tree topologies. Mol Biol Evol 13:964–969
24. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599. doi:10.1093/molbev/msm092
25. Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526
26. Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum likelihood approach. Mol Biol Evol 18:691–699
27. Yamasaki M, Kinashi H (2004) Two chimeric chromosomes of *Streptomyces coelicolor* A3(2) generated by single crossover of the wild-type chromosome and linear plasmid SCP1. J Bacteriol 186:6553–6559. doi:10.1128/JB.186.19.6553-6559.2004
28. Zucko J, Cullum J, Hranueli D, Long PF (2011) Evolutionary dynamics of modular polyketide synthases, with implications for protein design and engineering. J Antibiot 64:89–92. doi:10.1038/ja.2010.141