Evolution of fungal β-lactam biosynthesis gene clusters

Carlos García-Estrada1, Francisco Fierro1,2, and Juan-Francisco Martín1,2
1Institute of Biotecnology of León (INBIOtec), Av. Real, 1, 24006, León, Spain.
2Area of Microbiology, Departament of Molecular Biology, University of León, Campus de Vegazana s/n, 24071, León, Spain.
3Departamento de Biotecnología, Universidad Autónoma Metropolitana-Unidad Iztapalapa, Avda. San Rafael Atlixco 186, Col. Vicentina, 09340 México D.F.

Filamentous fungi are microorganisms of great biotechnological interest due to their ability to synthesize a variety of bioactive secondary metabolites including β-lactam antibiotics, such as penicillins and cephalosporins. Hydrophobic (with aromatic side chains) penicillins are only produced by fungi, mainly Penicillium chrysogenum and Aspergillus (Emericella) nidulans, whereas hydrophilic cephalosporins are produced by both fungi (e.g. Acremonium chrysogenum) and bacteria (e.g. Streptomyces clavuligerus). The genes encoding the enzymes responsible for the β-lactam antibiotic biosynthesis are clustered in the producing microorganisms and the availability of sequence information of fungal and bacterial genes has given rise to two main hypotheses about the evolutionary origin of these genes: (i) horizontal gene transfer (HGT) from bacteria to fungi and (ii) vertical descent. The strongest arguments, such as the absence of introns and some other features of some genes in the clusters, are in favour of horizontal gene transfer. The recent finding of paralogues of the penicillin biosynthetic gene penDE (also known as aatA) in the genomes of several filamentous fungi, which do not produce penicillin, provides new insights into the evolution of the β-lactam biosynthesis genes.

Keywords bioactive secondary metabolites; β-lactams; pathway evolution; horizontal gene transfer; gene cluster assembly

1. Introduction to β-lactam antibiotics

The discovery of β-lactam antibiotics is one of the most significant milestones of the human history. The family of β-lactam antibiotics stands out from the other families because of their high antibacterial activity and low toxicity and for this reason, they are among the most commonly prescribed drugs.

The fortuitous discovery of the antimicrobial activity generated by a fungus culture that contaminated a Petri-dish with Staphylococcus sp. by Sir Alexander Fleming in 1928, represented the starting point for modern antibiotic therapy. Fleming identified the mould responsible for the antibacterial effect as Penicillium notatum [1] and three years later, it was shown that the β-lactam antibiotic penicillin was the active compound inhibiting the bacterial growth [2]. From that moment, efforts were made by several scientists to produce a stable penicillin and develop a massive penicillin production method, initially in Oxford and after the beggining of World War II, in USA. Amongst these scientist, Ernst B. Chain and Howard W. Florey, together with Sir Alexander Fleming, were awarded with the Physiology and Medicine Nobel Prize in 1945. Penicillin was initially isolated from P. notatum cultures, but the low titers produced by this microorganism and the antibiotic demand derived from World War II, made the selection of new strains of paramount importance. An improvement was obtained after the isolation of Penicillium chrysogenum NRRL 1951 from an infected cantaloupe in a local market at Peoria IL, USA. This strain was more suitable than P. notatum for penicillin production in submerged cultures [3] and was subjected to rounds of classical mutagenesis during industrial improvement programs, which gave rise to the current penicillin high-producing strains. After the discovery in 1959 of the 6-aminopenicilanic acid (6-APA), which is a penicillin precursor, the synthesis of semisynthetic penicillins started up a new era of chemotherapy. Although the industrial production of penicillin is achieved with P. chrysogenum, other filamentous fungi such as Aspergillus nidulans (an ascomycete with sexual cycle) and Penicillium nalgiovense, are also able to produce this antibiotic.

The history of cephalosporins started in 1945, when the fungus Cephalosporium acremonium was isolated by Giuseppe Brotzu from the bay water at Cagliary (Italy) [4]. This fungus was renamed Acremonium chrysogenum because in Greek means “gold-producing branches”, whereas C. acremonium means “branches with head-like seeds”, which appeared to be less apt. It was at Oxford where this fungus was found to produce at least three types of antimicrobial compounds, which were isolated and identified [5]. The first compounds isolated in 1949 were members of the cephalosporin P complex and were later found to be tetracyclic triterpenes chemically related to helvolic acid (fumigacin). Cephalosporin P received this name because it was exclusively active against Gram-positive bacteria. Later the same year, a second compound initially named cephalosporin N, was found in culture filtrates from which cephalosporin P had been removed. This compound was active against Gram-negative and Gram-positive bacteria and was found to be a penicillin with a D-γ-aminoacidic side chain. Therefore, it was renamed penicillin N. Finally, in an experiment carried out in 1953 to determine the molecular weight of penicillin N, cephalosporin C was isolated. This antibiotic showed two interesting aspects since it was active against Gram-negative and Gram-positive bacteria and it was not hydrolysed by penicillinase. The latter was specially relevant due to the appearance of penicillin-resistant
bacteria. The main drawback of cephalosporin C was the weak antibacterial activity, but the isolation of overproducing mutants and the biosynthesis of semi-synthetic cephalosporins have solved this problem.

Antibiotics from the β-lactam family are not only produced by filamentous fungi. Gram-positive actinomycetes such as *Streptomyces clavuligerus* or *Amycolatopsis* (*Nocardia*) *lactandurans*, and Gram-negative bacteria such as *Lysobacter lactamgenus* also synthesize a variety of β-lactam structures including cephalosporins (mainly as intermediates of biosynthetic pathways), cephamycins, cephabacins, clavams, carbapenems and monobactams [6]. Cephamycins are 7-methoxy-cephalosporins active against penicillin-resistant bacteria. Cephabacins contain a formylamino group at C-7 and an oligopeptide side chain at C-3. The C-7 formilamino substituent of cephabacins and the C-7 metoxy group of cephamycins confers to these antibiotics their characteristic β-lactamase resistance.

2. Structure of β-lactam antibiotics

The β-lactam antibiotics, like many other secondary metabolites, have unusual chemical structures. As shown in Figure 1, all β-lactams contain a four-membered β-lactam ring closed by an amide bond and can be classified according to their chemical structure into five groups: penam, cephem, clavam, carbapenem and monolactam (monobactams and nocardicins). With exception of monolactams, which have only the β-lactam ring, the rest of β-lactam antibiotics consists of a bicyclic system. The structure of the second ring is used for their classification. They also contain side chains conferring hydrophilic or hydrophobic properties to these antibiotics [7].

![Classification of β-lactam antibiotics according to their chemical structure](image)

Penicillins contain a bicyclic ‘penam’ nucleus formed by fused β-lactam and sulphur-containing thiazolidine rings and an acyl side-chain such as phenylacetate, in the case of benzylpenicillin (penicillin G), or phenoxacetate in the case of phenoxymethylpenicillin (penicillin V), bound to the amino group at C-6. Cephalosporins contains the cephem nucleus, a six-membered dihydrothiazine ring fused to the β-lactam ring. Cephalosporin C has a D-α-aminoacidipyl side
chain attached to the C-7 amino group, which is identical to that of hydrophobic penicillin N but differs from those of hydrophilic penicillins. The activity of penicillins and cephalosporins as inhibitors of peptidoglycan biosynthesis in bacteria, make them specially interesting for clinical purposes. In the cephamycin family, the cephem nucleus contains, in addition to the α-aminoacyl side-chain, a methoxyl group at C-7. The family of cephabacins frequently contains a formyl group at C-7, and different peptides are attached to the C-3 carbon of the dihydrothiazine ring in the cephem nucleus. As indicated above, the groups attached to the C-7 of cephamycins and cephabacins are responsible for the insensitivity of these antibiotics to most β-lactamases [8].

The previous β-lactam compounds have a common mode of action, are synthesized from similar precursors and share several steps in their biosynthetic pathways. In addition to these classical β-lactam compounds, many non-conventional β-lactam structures have been discovered and characterized since 1970. As shown in Figure 1, these non-conventional β-lactams contain a β-lactam ring and they usually have a distinct bicyclic structure. The second ring in the molecule of clavulanic acid and other clavams is an oxazolidinic ring that includes oxygen instead of the sulfur atom occuring in classical β-lactams. The members of carbapenem and the olivanic acid family have a carbapenem ring containing a carbon atom instead of sulfur. Finally, monolactams (nocardicins and monobactams) contain a monocyclic structure (the β-lactam ring) and different side-chains. Non-conventional β-lactams are also inhibitors of peptidoglycan biosynthesis (monolactams), others are potent β-lactamase inhibitors with weak antibiotic activity (such as clavulanic acid), or have antifungal activity (some clavams). The precursors and biosynthetic pathways of non-conventional β-lactam antibiotics are different from those of classical β-lactams and the biosynthetic enzymes, genes, and gene organization are also different [9, 10]. In addition, non-conventional β-lactam antibiotics have never been found to be produced by eukaryotic cells, pointing to a different evolutionary origin of the genes encoding the biosynthetic enzymes.

3. Penicillin, cephalosporin and cephamycin biosynthesis

The biochemical pathways leading to classical β-lactam antibiotic biosynthesis have been largely studied and detailed information on the specific steps can be consulted in other review articles [11-15]. The biosynthesis of the β-lactam compounds involves sequential reactions including the formation of the penam nucleus from a linear tripeptide that is cyclized (early biosynthetic steps), ring expansion of the penam to the cepham nucleus (intermediate steps) and modification of the β-lactam nucleus (late decorating steps). The number of steps depends on the structure of the compound. As it can be seen in Fig. 2, the shortest pathway is that for penicillin biosynthesis and the more complex pathways occur in the biosynthesis of cephamycins and cephabacins.

The basic structure of all the classical β-lactam antibiotics originates from three amino acids: L-α-aminoacid, L-cysteine and L-valine. L-valine and L-cysteine are common amino acids but L-ω-aminoacidic is a non-proteinogenic amino acid formed by a specific pathway related to lysine biosynthesis. In fungi, ω-aminoacidic is an intermediate of the lysine biosynthesis pathway. In addition, lysine is catabolized to ω-aminoacidic acid in P. chrysogenum (i) by an ω-aminotransferase, encoded by the oat1 gene, which is induced by lysine [16], and (ii) by a reversal of the lysine biosynthesis pathway catalysed by the enzymes saccharopine dehydrogenase/saccharopine reductase [16, 17]. In β-lactam producing bacteria, lysine is converted into α-aminoacidic acid semialdehyde by lysine-6-aminoacid transferase (LAT), encoded by the lat gene. This semialdehyde cyclizes spontaneously to form piperidine-6-carboxylate, which is oxidized to α-aminoacidic acid by a piperidine-6-carboxylic acid dehydrogenase (P6C-DH), encoded by the pcd gene [18, 19].

Two early enzymatic steps are common to all the classical β-lactam producers, giving rise to the biosynthesis of isopenicillin N (IPN), which is the first compound in the pathway with antibiotic activity. The first enzyme of the pathway is the non-ribosomal peptide synthetase L-δ(α-aminoacidyl)-L-cysteinyl-D-valine (ACV) synthetase (ACVS), which is a very large multifunctional protein (Mr in the order of 420 kDa). This protein is encoded by a single structural 11-kb intronless gene named pcbAB or acvA, which is present in both fungal and bacterial penicillin and cephalosporin (and cephamycin) gene clusters (Fig. 3). The ACVS sequentially activates the three substrate amino acids with ATP to form α-aminoacyl-adenylates, binds them to the enzyme as thioesters, epimerizes the L-valine to D-valine, links together the three amino acids to form the tripeptide ACV and, finally, releases this tripeptide from the enzyme by the action of an internal thioesterase activity. ACVS consist of three well conserved domains that activate each of the three amino acids, respectively [20-22]. In the second step of the early biosynthetic stage, four hydrogen atoms from the ACV tripeptide are removed leading to the oxidative ring closure of the tripeptide and the formation of the bicyclic structure (penam nucleus) of IPN. This reaction is catalyzed by the IPN synthase (IPNS, cyclase), which is encoded by the intron-free pcbC (also known as ipn4) gene. IPNS is an intermolecular dioxygenase that require Fe²⁺, molecular oxygen and ascorbate. As indicated before, IPN constitutes the branch point of penicillin and cephalosporin biosynthesis.
The process of direct formation of the bicyclic structure of IPN differs from that of the other non-conventional β-lactam antibiotics, which first form the β-lactam ring and then, using a different enzyme, cyclize the monocyclic intermediate to form the second ring (oxazolidinic or carbapenem) present in the molecule, although it is probable that the two steps are coupled in vivo. Monolactam producers lack the ability to close the second ring of the nucleus.

In addition to the pcbAB (acvA) and pcbC (ipnA) genes common to filamentous fungi and bacteria, the fungi producing hydrophobic penicillins (i.e. *P. chrysogenum* and *A. nidulans*) contain a third gene in the penicillin cluster,
named *penDE* (*aatA*). This gene is not present in cephalosporin C or cephamycin-producing microorganisms. Unlike the other two genes in the cluster, the *penDE* gene contains three introns and encodes the peroxisomal acyl-CoA: IPN acyltransferase (IAT). This enzyme hydrolyses the α-aminoisobutyric side-chain of IPN and replaces it with a hydrophilic acyl molecule. These side chain precursors have to be previously activated by acyl-CoA ligases before they become substrates for the IAT [23]. A wide range of side chains may serve as substrates for this enzyme. Thus, natural penicillins, such as penicillin F (D3-hexenoic acid side chain) and K (octanoic acid as side chain) are synthesized under natural conditions. However, feeding the cultivation media with phenylacetic or phenoxyacetic acids, directs the biosynthesis mainly towards benzylpenicillin (penicillin G) or phenoxymethylpenicillin (penicillin V), respectively [24]. Recently, a novel penicillin gene named *aatB*, has been characterized initially in *A. nidulans*. This gene is not clustered with the rest of the penicillin genes (see below) and encodes a cytosolic protein, likely an acyl-CoA transferase, which plays a role in penicillin biosynthesis [25]. It is well known that the penicillin biosynthetic pathway is compartmentalized [15]. IPN biosynthesis takes place in the cytosol. However, both the side chain activation (catalysed by the phenylacetyl-CoA ligase) and the exchange of the side chain (catalysed by the IAT), occur inside microbodies (peroxisomes) [15, 23]. The distinct subcellular organization of penicillin biosynthesis implies transport of enzymes, precursors, intermediates and products through these compartments.

In the intermediate steps of cephalosporin biosynthesis, IPN is converted to its D-isomer (penicillin N), which is the precursor for antibiotics containing the cephem nucleus (i.e. cephalosporins and cephamycins). This conversion is carried out by a classical pyridoxal-phosphate-dependent epimerase in a single step in bacterial strains. This enzyme was first purified from actinomycetes [26], and the gene encoding its activity, *cefD*, was found to be located in the cephamycin gene cluster [27, 28]. However, epimerization of IPN in *A. chrysogenum* is encoded by a system constituted by two linked genes, *cefD1–cefD2*, which are located in the ‘early’ cephalosporin gene cluster (Fig. 3). The first gene, *cefD1*, has four introns and encodes a 71 kDa protein with similarity to fatty acid acyl-CoA synthetases. The second gene, *cefD2*, contains one intron and encodes a protein homologous to α-methyl-acyl-CoA racemases of eukaryotic origin. Disruption of either of these ORFs results in a lack of cephalosporin C production, loss of IPN epimerase activity and accumulation of IPN in the culture [29]. The proposed model for epimerization includes three biochemical steps: CefD1 converts IPN into isopenicillin N-CoA; then CefD2 isomerizes the compound into penicillin N-CoA, which seems to be released from the enzyme by the third enzyme, a thioesterase. There are also evidences for the compartmentalization of the cephalosporin biosynthesis pathway in *A. chrysogenum*. The two-component CefD1–CefD2 epimerization system seems to be located in microbodies, since canonical peroxisomal targeting sequences have been found in these two proteins [15]. This implies specific transport steps of precursors and intermediates across the peroxisomal membrane. This is supported by the finding of the MSF (major facilitator superfamily) transporter CefM, which is located in microbodies and is involved in the penicillin N secretion from the microbody lumen to the cytosol [30].

The following step in the cephalosporin/cephamycin pathway is the oxidative opening of the five-membered thiazolidine ring of penicillin N, giving rise to a six-membered dihydrothiazine ring upon reclosure. The ring expansion is catalysed by deacetoxycephalosporin C (DAOC) synthase (expandase) activity in both *A. chrysogenum* [31] and *N. lactamdurans* [32]. The next step of the pathway consists of the hydroxylation of the methyl group at C-3 of DAOC, forming deacetylcephalosporin C (DAC). As shown in Figures 2 and 3, both reactions are catalysed in *A. chrysogenum* by the same *cefEF*-encoded enzyme DAOC synthase (expandase)/DAC synthase (hydroxylase), whereas in *S. clavuligerus*, one enzyme for each reaction has been found: the DAOC synthase (encoded by the *cefE* gene) and the C-3 hydroxylase (encoded by *cefF* gene). Genes *cefE* and *cefF* encode proteins with about 70% identity in amino acids, which are 60% identical to the *cefEF*-encoded protein. In fact, these two enzymes have related molecular mechanisms, although each has retained approximately 10% of the residual activity of the other one. It is likely that these two genes are the consequence of a gene duplication event, giving rise to specialization to perform different, although mechanistically related, functions (expandase and hydroxylase).

The late (and final) step in cephalosporin C biosynthesis is the conversion of DAC to cephalosporin C by the DAC acetyltransferase, which uses acetyl-CoA as donor of the acetyl group. This enzyme encoded by the *cefG* gene has a Mr of 49 kDa and is evolutionary similar to O-acetyllhomoserine acetyltransferases [33]. As it is represented in Figure 3, the *cefG* gene contains two introns and is linked to the *cefEF* gene, but in the opposite orientation. In cephamycin-producing actinomycetes, a carbamoyl group is attached to DAC forming O-carbamoyl-DAC (ODAC). This reaction is catalysed by a carbamoyl transferase, which is encoded by the *cmcH* gene.

The last step of cephamycin biosynthesis consists of two sequential enzymatic reactions (a hydroxylation and the transfer of a methyl group to the hydroxyl present at C-7), which are carried out by a CmcI-CmcJ complex (encoded by the *cmcI* and *cmcJ* genes). This complex co-purifies by immunoaffinity chromatography and binds S-adenosylmethionine and deacetylcephalosporin C, in contrast to the isolated CmcI and CmcJ purified proteins, which do not have binding activity [34].

Concerning the cephabacin biosynthesis, previous studies were limited to a gene cluster of *L. lactamgenus* that encodes the enzymes responsible for the biosynthesis of the cephem nucleus (Fig. 3). The final steps of cephabacin biosynthesis have been deduced from the presence of several ORFs in the cephabacin gene cluster containing
nonribosomal peptide synthetase, a non-integrated thioesterase and polyketide synthase modules and a putative ABC transporter [35]. Their protein products are probably involved in the formation of the lateral chain in C-3 [36].

4. Organization and evolution of fungal β-lactam biosynthetic genes

Genes for β-lactam biosynthesis are clustered in all producer strains, either in bacterial or fungal microorganisms (Fig. 3). With minor differences between strains, the pcbC-pcbAB genes are always grouped and are located next to the penDE gene in penicillin producers. In A. chrysogenum, the “early” gene cluster, located on chromosome VII (4.6 Mb), contains the pcbAB and pcbC genes, encoding the enzymes for the first two steps of the pathway: cefD1 and cefD2 genes, responsible for the epimerization of IPN; and the cefT and cefM genes, which encode MSF proteins required for the transport of hydrophilic β-lactams and penicillin N, respectively [30, 37]. The “late” gene cluster, located on chromosome I (2.2 Mb), contains the cefEF and cefG genes, whose protein products are involved in the final steps of cephalosporin biosynthesis. The largest β-lactam clusters are those of cephamycin- and cephabacin-producing bacteria [8]. In S. clavuligerus, the cephamycin C gene cluster is adjacent to the clavulanic acid gene cluster. The entire supercluster of cephamycin C-clavulanic acid extends for about 50 kb. This organization of the biosynthetic genes of both antibiotics in a supercluster occurs also in other clavulanic-acid-producing strains, including Streptomyces jumonjinensis and Streptomyces katsurahamatus.

Gene clustering in fungi is an uncommon phenomenon, and it mainly occurs in pathways that are dispensable for viability, namely: catabolic pathways for utilization of unusual nutrients (e.g. ethanol, proline), and secondary metabolite biosynthetic pathways [39]. The clusters for the biosynthesis of β-lactam antibiotics have received much attention in the literature [4, 38, 40-42], and are very good examples of gene clustering and evolution. Secondary metabolite gene clusters usually include transport and regulatory genes along with the structural pathway genes. This is the pattern we observe when analyzing the cephamycin gene cluster of S. clavuligerus and A. lactamdurans (Fig. 3). In addition, resistance genes are present in these clusters, which is common in antibiotic clusters of prokayotes. Secondary metabolites are usually produced as families of compounds by a small group of related species and β-lactams are thus clearly an exception to this rule, as they are produced by members of taxa as phylogenetically distant as prokaryotic actinomycetes and eukaryotic ascomycetous fungi. The most widely accepted explanation for this is a horizontal gene transfer of the β-lactam genes between prokaryotes and fungi some time during evolution, which may have occurred approximately 370 million years ago [7, 11, 43]. The most important evidence supporting this hypothesis is that the fungal pcbAB and pcbC genes, responsible for the first two biosynthetic steps common to all β-lactam producers, present in their sequences some characteristics typical of actinomycetal (bacterial) genes and uncommon in fungal eukaryotic genes.
If we analyze the β-lactam clusters and pathways in prokaryotes and fungi, we can observe that the fungal pathways for penicillin and cephalosporin are shortened and divergent versions of the cephamycin pathway (Fig. 2), and in accordance with this, the fungal clusters contain fewer genes (Fig. 3). Two possible explanations may account for this finding: either prokaryotic clusters were only partially transferred to fungi, or some bacterial genes were eventually lost during evolution in fungi and only pcbAB and pcbC, plus cefEF in A. chrysogenum, became a stable part of the genomes of β-lactam producing fungi.

Another intriguing aspect of the penicillin and cephalosporin gene clusters is the absence of pathway-specific regulatory genes. This is not common, as fungal secondary metabolite clusters sometimes contain one or more genes encoding transcription factors that regulate expression of the genes in the clusters; for instance lovE in the A. terreus lovastatin gene cluster [44], Trif in the Fusarium sporotrichioides trichotheccene gene cluster and aflR in the A. nidulans and other Aspergillus spp. sterigmatocystin/aflatoxin gene clusters [45]. Nevertheless, regulatory genes are present in the cephamycin gene clusters of actinomycetes [6], and thus we may invoke one of the two above-mentioned possibilities to explain their absence in fungi. Instead of a specific regulator, other wide domain transcription factors, encoded by genes outside the clusters, were recruited by the β-lactam producing fungi to regulate penicillin and cephalosporin gene expression, such as CreA/Cre1, which mediates carbon source catabolic repression, PaeC, responsible for pH regulation, or AnCF, a CCAAT-binding complex [12, 24, 46]. The absence of penicillin and cephalosporin specific regulators is another evidence supporting the hypothesis of the horizontal gene transfer of β-lactam genes; in other secondary metabolite gene clusters of fungi, regulatory genes evolved together with structural genes in the development of a particular pathway and were vertically transferred to the progeny, whereas in β-lactam producing fungi evolution took place differently, incorporating genes of bacterial origin that were expressed under the control of trans-acting transcription factors to regulate them. The horizontal transfer of bacterial regulatory genes together with the β-lactam biosynthetic genes may have occurred, but most likely, they were lost because they were not functional in fungi.

As mentioned above, the penicillin and cephalosporin pathways are shortened and divergent versions of the cephamycin pathway in actinomycetes. Fungi from the Penicillium/Aspergillus lineage received (or conserved) from bacteria only the pcbAB and pcbC genes, developing subsequently a new pathway by incorporating a gene of eukaryotic origin, penDE, to synthesize the powerful antibiotic benzylpenicillin. Most likely, the penDE gene ancestor and the pcbAB-pcbC gene pair were originally separated in the genome, but eventually they became clustered, and conserved this organization after the split in two separated genera. The penicillin gene cluster was lost in many species from both genera, and conserved in species such as A. nidulans (in chromosome VI, [47]), P. chrysogenum (chromosome I, [48]), P. griseofulvum, P. nalgiovense or partially in P. verrucosum [49] (see below). Similarly, an A. chrysogenum ancestor received (or conserved) from bacteria the pcbAB, pcbC and cefEF genes, and subsequently developed a new pathway to produce cephalosporin, by recruiting the genes of eukaryotic origin cefG, which became clustered with the cefEF gene in the so-called “late” cephalosporin cluster (chromosome I, [41]), and cefD1 and cefD2, which became clustered with pcbAB and pcbC in the “early” cluster (chromosome VII, [41]). The case of cefD1 and cefD2 is particularly interesting, as they encode enzymes that catalyze the isomerization of IPN into penicillin N. As indicated before, this step also occurs in bacteria but it is catalyzed by a different enzymatic system, i.e, there was no transfer of the bacterial cefD gene to fungi, instead fungi independently developed a new system to isomerize isopenicillin N, whose responsible genes, cefD1 and cefD2, became clustered with the pcbAB and pcbC genes. Other genes of eukaryotic origin, cefT and cefM, encoding cephalosporin and intracellular precursor transporters respectively, are also present in the A. chrysogenum “early” cluster.

Why are secondary metabolite genes clustered in fungi whereas genes for many other biosynthetic pathways are not?. This intriguing question has received attention for a long time and different explanations have been proposed. When clustering of fungal β-lactam genes was first discovered, it was hypothesized to be a reflection of clustering in bacteria and a mere consequence of the horizontal transfer of genes. However, later studies on other fungal secondary metabolite clusters showed that clustering was a general phenomenon. In addition, the way many genes dispersed in the genome, or the CCAATT-binding complex AnCF, which controls expression of thepcbAB-pcbC, plus cefEF in A. chrysogenum, became a stable part of the genomes of β-lactam producing fungi.

The recent discovery of the secondary metabolism global regulator LaeA has provided important clues to explain the reasons for secondary metabolite gene clustering in fungi. LaeA was first described in A. nidulans; ΔlaeA mutants showed reduced production of sterigmatocystin, penicillin and other secondary metabolites [50]. In P. chrysogenum, laeA knock-down mutants are affected in penicillin and pigment production as well as in sporulation [51]. LaeA has been shown to act at the transcriptional level, controlling expression of secondary metabolism clustered genes. It shows strong similarity to methyltransferase enzymes, for which it is believed to act epigenetically, causing a change from heterochromatic to euchromatic state, thus allowing coordinated expression of all genes in a cluster. This hypothesis is further supported by the observation that in A. nidulans, the exogenous primary metabolism gene argB inserted in the
sterigmatocystin cluster was silenced in a ΔlaeA background, but the expression of genes surrounding the cluster was unaffected [52].

Subsequently, additional examples of epigenetic regulation of whole secondary metabolite gene clusters have been found. In *A. nidulans*, histone deacetylase HdaA negatively regulates transcription of genes in the sterigmatocystin and penicillin biosynthetic clusters, both situated in telomere proximal regions, but not in the terrquinone A cluster, which is situated far from telomeric regions [53]. Subtelomeric regions are frequently associated with gene silencing induced by histone deacetylation. HdaA and LaeA seem to act independently in the regulation of secondary metabolite clusters; LaeA also controls terrquinone A gene expression, which indicates that its activity is independent from the position of the clusters with respect to the telomere.

In summary, secondary metabolite gene clustering in fungi seems to have important functional basis, being required for a coordinated expression of all the genes by global or local chromatin remodeling mechanisms, dependent on the function of the global secondary metabolite regulator LaeA and other chromatin remodeling enzymes.

5. Evolution of the penicillin biosynthetic gene *penDE (aatA)*

Unlike the *pcbAB (acvA)* and *pcbC (ipnA)* genes, the *penDE (aatA)* gene contains three introns, and several common features of this gene indicate that it was derived by fusion of two eukaryotic DNA fragments, one of them containing introns [43], and was recruited during evolution by fusion to the *pcbAB-pcbC* gene cluster from an ancestral fungal gene [7, 38]. The origin of the fungal *penDE (aatA)* gene has remained obscure because no known close eukaryotic or prokaryotic analogous had been identified. However, *in silico* analysis of the genomes of several ascomyces (including *A. nidulans* and *P. chrysogenum*) has recently allowed the identification of a putative gene paralogue of the *penDE (aatA)* gene that may help to clarify the evolution of the gene cluster. This novel gene has been designated *aatB* in *A. nidulans* [25] or *ial* in *P. chrysogenum* [54]. However, several differences have suggested that these two genes might have undergone modifications during evolution (see below).

The *aatB* gene of *A. nidulans* shows an expression profile similar to that of the *aatA* gene. Both genes contain three introns (four exons), but unlike the IAT encoded by the *aatA* gene, which includes a peroxisomal targeting sequence at the C-terminal end and is targeted to these organelles, the protein encoded by the *aatB* gene lacks this motif and is located on the cytosol. The role of the *aatB* gene in penicillin biosynthesis has been confirmed in *aatA*-disrupted mutants, proving that the protein encoded by the *aatB* gene was capable of partially replacing the IAT (encoded by the *aatA* gene) activity in those mutants. Another interesting finding was the fact that the same transcription factors AnCF and AnBH1 bind and regulate the promoter region of these two genes *in vitro* [25]. These authors suggested that because of the identical exon distribution and sequence similarity, the *aatA* and *aatB* genes are descendants of a single ancestral gene that became duplicated during evolution and that the *aatB* gene represents a paralogue of *aatA*. In this model, one of the counterparts was recruited to the penicillin gene cluster becoming the *aatA* gene.

The *ial* gene of *P. chrysogenum* contains two introns (three exons) and it also encodes a protein that lacks the canonical peroxisomal targeting sequence. Unlike the *A. nidulans* *aatB* gene, the *ial* gene is expressed very poorly or not expressed at all in several *P. chrysogenum* strains and generation of *ial* null and overexpressing mutants do not affect penicillin production. *In vitro* experiments carried out with the protein encoded by the *ial* gene revealed that despite the proper structure, folding and processing events, the protein lacked IPN acyltransferase activity [54]. Therefore, *aatB* and *ial* genes appear to differ in function.

The availability of the genome of several ascomycetes has revealed the presence of *ial* (*aatB*) gene homologues in penicillin and non-penicillin producing fungi, whereas the *penDE (aatA)* gene homologues are only found in penicillin-producing fungi, such as *A. nidulans* and *A. oryzae*. This might indicate that during evolution, a single ancestral gene was duplicated, giving rise to the *penDE (aatA)* gene and its paralogue, the *ial* (or *aatB*) gene (initially encoding a NTL amidohydrolase not active in *P. chrysogenum* and with low activity in *A. nidulans*). On the other hand, the *P. chrysogenum* IAL and related proteins in other fungi form a separate evolution clade from IATs (Fig. 4), indicating that they might have evolved separately. This hypothesis is supported by the presence of duplicated genes encoding putatives IAT and IAL homologues in *A. oryzae*, which also contains the penicillin gene cluster. From those ascomycetes containing this cluster, only *A. nidulans* has an IAL homologue (GenBank: XP_664379) more closely related to IATs, a fact that may explain the presence of penicillin biosynthetic activity in this protein. Genes encoding IATs in *P. chrysogenum*, *A. nidulans* and *A. oryzae* contain three introns, thus differing from those genes encoding IAL and IAL-homologues (with exception of *A. terreus ial* gene homologue). Only the *aatB* gene encoding the *A. nidulans* IAL homologue and the *A. terreus aatB* gene homologue (GenBank: XP_001213312), contain three introns (Fig. 4). This suggests that alternatively, *ial* and *ial* gene homologues might have had a different origin from the IAT-encoding genes (*penDE or aatA* genes), thus encoding proteins with a different function as it was confirmed by the lack of penicillin biosynthetic activity of the *P. chrysogenum* IAL. With this hypothesis, only the *aatB* gene from *A. nidulans* would be a real paralogue of the IAT-encoding gene (*aatA*) formed by gene duplication from a common ancestor. This is supported by the presence of penicillin forming activity of the *aabB*-encoded IAL homologue and by the presence of the same transcription factors binding to the promoter regions of these two genes [25]. The question of when the
duplication of *ial* and *aatB* genes occurred forming the penicillin biosynthesis genes *penDE* and *aatA* cannot be satisfactorily answered at the current moment.

If there was a common ancestor for the *ial* and *penDE* genes, most of the Ascomycota fungi initially had the potential capacity to perform the acyltransferase reaction. However, only a few of them, like *A. nidulans* and *P. chrysogenum*, were able to develop, during evolution, the *penDE* (*aatA*) gene encoding the highly functional IAT enzyme. The *penDE* (*aatA*) gene was linked to the first two genes (of bacterial origin) of the penicillin pathway, which endowed these microorganisms with an important ecological advantage because of the ability to generate aromatic penicillins. It is likely that the *de novo* formation of this cluster occurred in a common ancestor of the genera Penicillium and Aspergillus, since the *pen* cluster is present in several species of those genera. However, not all genomes of the aspergilli contain the *pen* cluster; e.g., *A. fumigatus* lacks it, although it contains the *ial* gene. This indicates that the *pen* cluster might have been horizontally transferred only to some species of the genus, or alternatively, the primitive pen cluster might have been lost during subsequent evolution. A model of the evolution of the *penDE* (*aatA*) and *ial* (*aatB*) genes is represented in Figure 5.
Fig. 5. Schematic representation of an evolutionary model proposed for the penDE (aatA) and ial (aatB) genes. In this model, the duplication of the ancestral gene that gave rise to the penDE (aatA) and ial (aatB) genes happened after the splitting of the genera Aspergillus and Penicillium and the penDE (aatA) gene was horizontally transferred to particular species with the other penicillin biosynthetic genes.

Acknowledgements. The support by grants of the Eupean Union (Eurofung QLRT-1999-00729 and Eurofungbase), Agencia de Inversiones y Servicios de Castilla y León (Proyecto Genérico de Desarrollo Tecnológico 2008) and the Ministry of Industry of Spain (PROFIT: FIT-010000-2007-74) is gratefully acknowledged. Authors also wish to thank B. Martín, J. Merino, A. Casenave and A. Mulero (INBIOTEC) for the excellent technical assistance.

References

Liras P, Martin JF. Gene clusters for β-lactam antibiotics and control of their expression: why have clusters evolved, and from where did they originate?. *International Microbiology.* 2006;9:9-19.


Martín de Valmaseda EM, Campoy S, Naranjo L, Casqueiro J, Martin JF. Lysine is catabolized to α-aminoadipic acid in *Penicillium chrysogenum* by an omega-aminotransferase and to saccharopine by a lysine 2-ketoglutarate reductase. *Molecular Genetics and Genomics.* 2005;274:272-82.


[48] Fierro F, Gutiérrez S, Diez B, Martin JF. Resolution of four large chromosomes in penicillin producing filamentous fungi: The penicillin gene cluster is located on chromosome II (9.6 Mb) in *Penicillium notatum* and chromosome I (10.4 Mb) in *Penicillium chrysogenum.* *Molecular and General Genetics.* 1993;241:573-578.


