Bioorganic Chemistry 44 (2012) 1–7

Contents lists available at SciVerse ScienceDirect

Bioorganic Chemistry

journal homepage: www.elsevier.com/locate/bioorg

Highly Important Paper

IMPORTANT

# Insights into fluorometabolite biosynthesis in *Streptomyces cattleya* DSM46488 through genome sequence and knockout mutants

Chunhua Zhao [a,1], Peng Li [b,1], Zixin Deng [b], Hong-Yu Ou [b,*], Ryan P. McGlinchey [a], David O'Hagan [a,*]

[a] School of Chemistry and Centre for Biomolecular Sciences, University of St. Andrews, North Haugh, St. Andrews, Fife KY16 9ST, UK
[b] State Key Laboratory of Microbial Metabolism and School of Life Sciences & Biotechnology, Shanghai Jiaotong University, Shanghai 200030, China

ARTICLE INFO

*Article history:*
Received 29 May 2012
Available online 16 July 2012

*Keywords:*
Fluorometabolites
Fluoroacetate
4-Fluorothreonine
*Streptomyces cattleya*
Biosynthesis
Genome sequence

ABSTRACT

*Streptomyces cattleya* DSM 46488 is unusual in its ability to biosynthesise fluorine containing natural products, where it can produce fluoroacetate and 4-fluorothreonine. The individual enzymes involved in fluorometabolite biosynthesis have already been demonstrated in *in vitro* investigations. Candidate genes for the individual biosynthetic steps were located from recent genome sequences. *In vivo* inactivation of individual genes including those encoding the *S*-adenosyl-L-methionine:fluoride adenosyltransferase (fluorinase, SCATT_41540), 5′-fluoro-5′-deoxyadenosine phosphorylase (SCATT_41550), fluoroacetyl-CoA thioesterase (SCATT_41470), 5-fluoro-5-deoxyribose-1-phosphate isomerase (SCATT_20080) and a 4-fluorothreonine acetaldehyde transaldolase (SCATT_p11780) confirm that they are essential for fluorometabolite production. Notably gene disruption of the transaldolase (SCATT_p11780) resulted in a mutant which could produce fluoroacetate but was blocked in its ability to biosynthesise 4-fluorothreonine, revealing a branchpoint role for the PLP-transaldolase.



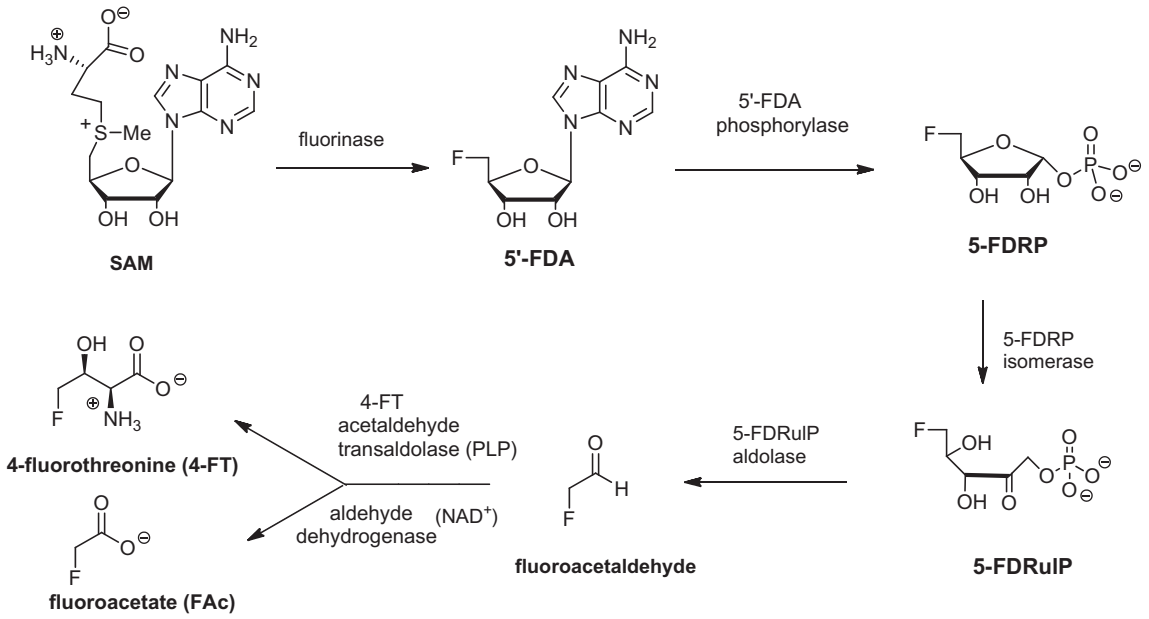
## 1. Introduction

Fluorine is the thirteenth most abundant element in the earth's crust, and the most abundant halogen, however fluorinated natural products are extremely rare [1]. By contrast fluorochemicals are of huge commercial importance, and a large organofluorine industry underpins the development of new products for the pharmaceuticals, agrochemicals and organic materials sectors [2]. These products are all prepared by chemical methods and presently the chemicals industry does not have any biotechnology approaches for the generation of fluorinated intermediates or fine chemicals. The mechanisms of biohalogenation, particularly for bromine and chlorine, have been widely discussed [3]. Whilst a halide ion may be used directly as a nucleophile, more commonly halogenases utilise oxidative strategies removing either one or two electrons from the halide ions to generate either a halide radical or halonium ion equivalent. The actinomycete *Streptomyces cattleya* has the rare ability to biosynthesise fluorometabolites, and can biosynthesise fluoroacetate (FAc) and 4-fluorothreonine (4-FT) as illustrated in Scheme 1 [4]. The fluorinase enzyme which mediates C–F bond formation in this organism [5] catalyses the nucleophilic attack of a fluoride ion to *S*-adenosyl-L-methionine (SAM) to generate 5′-fluoro-5′-deoxyadenosine (5′-FDA) [6]. 5′-FDA is then converted to the toxin FAc and the antibiotic 4-FT. *S. cattleya* also has the capacity to biosynthesise other antibiotics such as thienamycin [7] and cephamycin C [8].

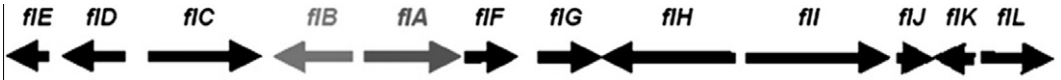
Previous *in vitro* investigations have established that five enzymes are involved in the biosynthesis of 4-FT (Fig. 1): Firstly, the fluorinase (*flA*) catalyses the reaction of *S*-adenosyl-L-methionine (SAM) and a fluoride ion to generate 5′-FDA and L-methionine [5]. Then an adenosine phosphorylase (*flB*) mediates nucleophilic phosphate attack upon 5′-FDA to remove the adenosine base and generate 5-FDRP (5-fluoro-5-deoxy-D-ribose-1-phosphate) [9]. The next enzyme catalyses a ring opening isomerisation which converts 5-FDRP to 5-fluoro-5-deoxy-ribulose-1-phosphate (5-FDRulP) [10]. 5-FDRulP is the product of an aldolase, and thus a reverse aldol reaction generates fluoroacetaldehyde, which is the last common intermediate before a branch point in the metabolism. Either a $NADH^+$-dependent dehydrogenase oxidises fluoroacetaldehyde to fluoroacetate [11] or a pyridoxal phosphate (PLP)-dependent transaldolase catalyses the cross condensation reaction of fluoroacetaldehyde and L-threonine to yield 4-FT and acetaldehyde [12]. Five proteins required for 4-FT biosynthesis have been obtained directly by over-expression in *Escherichia coli* of the relevant genes from *S. cattleya* (fluorinase, phosphorylase, isomerase and transaldolase) or in the case of the aldolase, by acquiring a homologous surrogate enzyme from a related streptomycete. When the combined over-expressed enzymes were placed *in vitro* and the buffer supplemented with fluoride ion, SAM, L-threonine and PLP, a complete synthesis of 4-FT was achieved in an NMR tube [13].

* Corresponding authors. Fax: +86 21 62932418 (H.-Y. Ou), +44 116 2525030 (D. O'Hagan).
*E-mail addresses:* hyou@sjtu.edu.cn (H.-Y. Ou), do1@st-andrews.ac.uk (D. O'Hagan).
[1] These authors contribute equally to this study.

0045-2068/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bioorg.2012.06.002




**Scheme 1.** Biosynthetic pathway to the fluorometabolites, fluoroacetate (FAc) and 4-fluorothreonine (4-FT) in *S. cattleya*.




| ORF | Start/Stop (bp) | Length (aa) | Function/Homology |
|---|---|---|---|
| E | 130-795c | 222 | DNA binding regulatory protein |
| D | 857-1504c | 216 | Dehalogenase/Phosphatase |
| C | 1845-3036 | 397 | MFS permease |
| B | 3057-3953c | 299 | 5'-FDA phosphorylase |
| A | 4173-5069 | 299 | 5'-FDA synthase |
| F | 5197-5751 | 185 | DNA binding regulatory protein |
| G | 5951-6652 | 234 | DNA binding regulatory protein |
| H | 6652-8052c | 467 | $Na^+/H^+$ antiporter |
| I | 8314-9780 | 489 | Homocysteine lyase |
| J | 9803-10195 | 131 | DNA binding protein |
| K | 10592-10176c | 139 | Thioesterase/acyltransferase |
| L | 10700-11374 | 225 | DNA binding regulatory protein |

**Fig. 1.** The 'Spencer' 12 kb gene cluster of 2006 [16] which identifed the fluorinase gene (*flA*) and the adjacent PNP (*flB*) gene. However the relationship of the surrounding genes to the biosynthesis of the fluorometabolites is not clear.

Additionally the identity of the intermediate metabolites 5′-FDA [14], 5-FDRP [9], 5-FDRulP [10] and fluoroacetaldehyde [15] have been established either in whole cell or cell-free studies. The first two genes of the fluorometabolite pathway, encoding the fluorinase (*flA*) and the adenosine phosphorylase (*flB*), lie adjacent to each other on the chromosome. This relationship was identified in 2006, when a 12 kb fragment of chromosomal *S. cattleya* DNA was sequenced on either side of the fluorinase gene (*flA*) [16]. *FlA* and *flB* were recently shown to share high homology to a related pair of genes in the salinosporamide A biosynthetic pathway (*salL* and *salT*), a metabolite elaborated by the marine actinomycete *Salinispora tropica* [17]. *SalL* expresses a chlorinase, an enzyme that catalyses a similar nucleophilic reaction with SAM to the fluorinase, but using chloride ion rather than fluoride ion, to generate 5′-chloro-5′-deoxyadenosine (5′-ClDA) [11]. When the fluorinase gene was used to replace the chlorinase gene in *S. tropica*, a hybrid antibiotic fluorosalinosporamide was elaborated [18]. This demonstrated that the fluorinase could be used to complement a different biosynthetic pathway and highlighted the future biotechnological potential for novel fluorinated bioactive production by organism engineering.

Twelve genes were identified in the 12 kb fragment of *S. cattleya*, with *flA* (fluorinase) and *flB* (PNP) situated adjacent to each other, but notably no other biosynthetic genes were obvious [16]. Of these genes, *flK* has subsequently been confirmed as a fluoroacetyl-CoA thioesterase [19]. *FlK* appears to confers self-resistance of the host to the toxin fluoroacetate by hydrolysing any fluoroacetyl-CoA formed and therefore suppresses accumulation of fluoroacetyl-CoA which is converted to fluorocitrate by citrate synthase. Fluorocitrate is a potent inhibitor of aconitase and citrate transporting proteins and shuts down the Krebs cycle [20]. The gene *flH* has homology to a $Na^+/H^+$ antiporter. Its proximity to *flA* suggested that

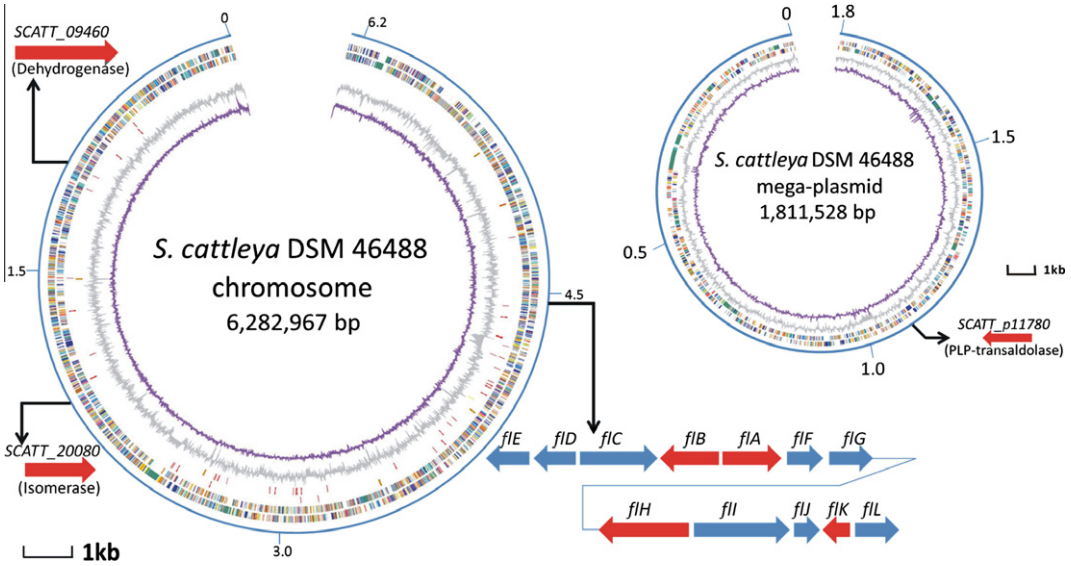



**Fig. 2.** Genome sequencing of *S. cattleya* DSM 46488. The genome consists of a linear chromosome (6.3 Mb, 5222 genes) and a linear mega-plasmid (1.8 Mb, 1747 genes). The genes involved in the biosynthesis of fluorometabolites are marked. The key genes knocked out in this study are highlighted in red. The data complement an earlier genome deposition [21].

it may be involved in fluoride ion uptake into the cell, with neutral HF crossing the membrane, and then the antiporter exchanging $H^+$ for $Na^+$ ions, and securing NaF within the cell. The other biosynthetic genes of the fluorometabolite pathway, such as the 5-FDRP isomerase, 5-FDRulP aldolase and the PLP dependent transaldolase were not located on this fragment, and thus further genome sequencing was desirable to identify their locations. In general it is emerging that the genes of most biosynthetic pathways in *Streptomyces* are clustered. The fluorometabolite pathway has the apparent complexity that only two biosynthesis genes lie adjacent to each other. A full genome sequence of *S. cattleya* NRRL 8057 [21], carried out during our study identified a genome consisting of a large (6.28 Mb) linear chromosome and a smaller (1.8 Mb) linear mega-plasmid. That analysis indicated that the remaining genes encoding enzymes for the biosynthesis were distributed across the large chromosome, and with the PLP-transaldolase located on the mega-plasmid. We have also sequenced the genome of *S. cattleya* (DSM 46488), to explore gene organisation and as a prelude to gene knockout experiments.

## 2. Results and discussion

### 2.1. Genome sequencing of S. cattleya DSM 46488

The *S. cattleya* strain DSM 46488 genome was sequenced and confirmed an almost identical genome to that of *S. cattleya* NRRL 8057 submitted recently [20]. This genome contains two replicons. Specifically a linear chromosome of 6,282,967 bp (72.94% G+C content) and a linear mega-plasmid of 1,812,548 bp (73.27% G+C content). The general sequence features of the two replicons are shown in Fig. 2 and in Table 1 of the Supplementary information. The chromosome coded for 5822 putative proteins, 41.1% of which were annotated as hypothetical. Six rRNA operons (16S-23S-5S), 63 tRNA and 1 tmRNA genes were also identified on the chromosome. Interestingly, the mega-plasmid pSCATT (1.8 M) is one of the largest linear plasmids to be sequenced. There was neither an rRNA, tRNA or a tmRNA gene identified on the mega-plasmid; 1747 protein-coding genes were found, 38.9% of which coded for hypothetical proteins.

The nucleoside sequence and annotated proteins of the chromosome and megaplasmid pSCATT have been deposited in GenBank CP003219 (SCATT_00010–SCATT_58220 representing a total of 5822 proteins) and CP003229 (SCATT_p00010–SCATT_p17470 representing a total of 1747 proteins) respectively. The data complement that of the genome of *S. cattleya* NRRL8057 deposited in the EMBL (accession numbers FQ859185 and FQ859184) [21].

The previously sequenced 12-kb fluorometabolite biosynthesis gene cluster, which contains *flA* and *flB*, is localised on the region SCATT_41460–SCATT_41590 of the DSM 46488 chromosome. This also locates the fluoroacetyl-CoA hydrolyse (FlK, SCATT_41470) [19], fluorinase (FlA, SCATT_41540) and the 5′-fluoro-5′-deoxyadenosine phosphorylase (FlB, SCATT_41560) [16]. As previously noted [19], BLASTp searches highlight that the other four genes involved in the biosynthesis of fluorometabolites are found to be dispersed over the chromosome and the mega-plasmid, distant from *flA*. The 5-FDRP isomerase (SCATT_20080), the acetaldehyde dehydrogenase (SCATT_09460) and four aldolases (SCATT_10350, SCATT_33160, SCATT_p05770, SCATT_p01080), one of which is anticipated to be the relevant fluorometabolite aldolase, were located on the large chromosome. The gene encoding the 4-FT acetaldehyde transaldolase (SCATT_p11780) is located on the mega-plasmid [21].

### 2.2. Generation of mutants by inactivation of genes clustered with flA (fluorinase)

Genome sequencing of *S. cattleya* has revealed a cluster of three genes (SCATT_41630–SCATT_41650) located seven genes upstream of *flA*. These were a homoserine kinase, a threonine synthase and a homoserine dehydrogenase, a complement required for L-threonine biosynthesis. L-Threonine is a required substrate for 4-FT acetaldehyde transaldolase involved in the last step in 4-FT biosynthesis and the location of these genes proximate to *flA*/*flB* suggests a role in providing an important precursor for 4-FT biosynthesis.

#### 2.2.1. Inactivation of flA (SCATT_41540, fluorinase)

*In vitro* studies have demonstrated that *flA* encodes the fluorinase, a 299 amino acid protein which combines fluoride ion ($F^-$) and SAM to generate 5′-FDA and L-methionine [5]. In order to establish if exogenous 5′-FDA supplementation would recover fluorometabolite production of a *flA* knockout mutant of *S. cattleya*, inactivation

of *flA* was carried. This experiment was designed to explore if the products of *flA* and *flB* can operate independently, despite being located adjacent to each other on the chromosome. Accordingly a 1386 bp apramycin resistance cassette from pIJ773 (see Section 4) was inserted into an internal 900 bp fragment of *flA* (Fig. 3). Sequencing validated that the apramycin resistance cassette had inserted into *flA* at the expected locus (see Supplementary data S2). The *flA* knockout mutant was cultured in fluoride enriched media and as expected it was unable to produce fluorometabolites. However when 5′-FDA (2 mM) was added into a 5 d culture broth, then this restored fluorometabolite production in the *flA* mutant of *S. cattleya* after a 48 h incubation, consistent with the metabolite supplementation by-passing the *flA* knockout and indicates that the PNP enzyme is not reliant on a functional fluorinase.

*2.2.2. Inactivation of flK (SCATT_41470, fluoroacetate CoA hydrolase)*

*FlK* has been over-expressed in *E. coli* and shown to be a 139 amino acid thioesterase able to catalyse the hydrolysis of fluoroacetyl-CoA to fluoroacetate [16]. The enzyme displays a remarkable selectivity in that it is unable to catalyse hydrolysis of acetyl-CoA. It has been suggested that the gene confers resistance to fluoroacetate toxicity. Indeed cloning of the gene into *E. coli* does increase the tolerance of the cloned organism to wild type to fluoroacetate [19]. To further explore the role of this thioesterase the *flK* gene was inactivated by inserting the 1384 bp apramycin resistance cassette into an internal fragment of *flK*. Subsequent sequencing of PCR amplified DNA showed that 1.4 kb of DNA was inserted into the expected locus on the chromosome (see Supplementary data S2). This *flK* knockout mutant was unable to generate any fluorometabolites on incubation with fluoride ions, thus there may be some regulatory aspect associated with *flK* which shuts down the pathway when the resistance gene is not expressed. In order to assess if the mutant was now susceptible to fluoroacetate toxicity, it was grown in a media supplemented with fluoroacetate, fluoroacetaldehyde and ethyl fluoroacetate up to concentrations of 20 mM. In each case the cultures grew well, similar to wild type cultures, perhaps suggesting that *S. cattleya* is unable to take up exogenous and toxic fluoroacetate.

*2.2.3. Inactivation of flB (SCATT_41550, 5′-fluoro-5′-deoxyadenosine phosphorylase)*

The *flB* gene, which encodes the second enzyme, a 5′-fluoro-5′-deoxyadenosine phosphorylase (PNP), on the biosynthetic pathway, was inactivated. Accordingly the apramycin resistance cassette was inserted into the *flB* gene as described in Section 4. As expected, after culturing and incubation with fluoride (2 mM) the *flB* gene had lost its ability to produce both FAc and 4-FT. The inactivation was validated by PCR and subsequent DNA sequencing (Supplementary data S2).

*2.2.4. Inactivation of flH (SCATT_41500 putative $Na^+/H^+$ antiporter)*

The *flH* gene has homology to other bacterial $Na^+/H^+$ antiporters. Its location close to the fluorinase gene (*flA*) indicated that it may play a role in managing cellular fluoride ion uptake. In order to explore the influence of the *flH* gene on fluorometabolite biosynthesis, inactivation was carried out as described in Section 4. A 1401 bp fragment internal to *flH* was substituted by the apramycin resistance cassette. Appropriate primers were used to amplify chromosomal DNA of the *flH* mutant as well as that of wild type *S. cattleya*. DNA with the predicted insert size (1.5 kb) was amplified from the mutant and subsequent sequencing showed that the apramycin resistance cassette had indeed inserted into the expected locus of the *S. cattleya* chromosome (see Supporting information S2).

Interestingly this mutant retained full competence in fluorometabolite biosynthesis generating similar levels of FAc and 4-FT (Fig. 2B), to that of wild type *S. cattleya*. Consequently *flH* is not essential for fluorometabolite production, and does not have a significant role in regulating fluoride ion uptake.

*2.3. Generation of mutants by inactivation of genes not clustered with flA*

Genome sequencing [21] has enabled the identification of genes, beyond the Spencer cluster, which may be associated with fluorometabolite biosynthesis. These were the 5-FDRP isomerase which converts 5-FDRP to 5-FDRulP, the aldolase which mediates a retro-aldol reaction of 5-FDRulP to generate fluoroacetaldehyde and then finally the PLP dependent 4-FT acetaldehyde transaldolase involved in the last step of 4-FT biosynthesis. Candidate genes were identified for these enzymes by a combination of prior purification and amino acid sequencing of the enzymes (PLP-transaldolase) or by homology to genes in *Streptomyces coelicolor* after sequencing [20] (5-FDRP isomerase, 5-FDRulP aldolase).

*2.3.1. Inactivation of a putative 5-FDRP isomerase (SCATT_20080)*

A 5-FDRP isomerase gene (1161 bp) from *S. cattleya* has previously been cloned into *E. coli* and over-expressed, and then shown to be able to convert 5-FDRP to 5-DRulP. The gene was identified by its homology to a thiomethylribose phosphate isomerase from *S. coelicolor* [13]. Full genome sequencing re-identified this gene (SCATT_20080) and we set out to establish if it is the sole isomerase in the *S. cattleya* chromosome competent to mediate the conversion to 5-DRulP during fluorometabolite biosynthesis. Using this isomerase gene as a DNA probe, a fosmid (29C1) was identified from a genome library of *S. cattleya*, which harboured the gene. Then, using the REDIRECT system [27] as described in Section 4, an apramycin resistance and chloramphenicol sensitive strain of *S. cattleya* was constructed in which the apramycin resistance cassette was inserted into an 1161 bp fragment of the isomerase gene. Insertion was validated by PCR and subsequent sequencing (see Supporting information S2). This proved to be a healthy organism when grown under standard conditions however the mutant was unable to biosynthesise FAc or 4-FT, suggesting that this isomerase is required for the isomerisation of 5-FDRP to 5-DRulP (Fig. 1).

*2.3.2. Four putative aldolases*

The fluorometabolite pathway requires an aldolase to catalyse the cleavage of 5-DRulP to generate fluoroacetaldehyde [8]. To date identification of the enzyme or gene for this aldolase has been elusive. In the *in vitro* cell free reconstruction of 4-FT biosynthesis, an aldolase from *S. coelicolor* (SCO1844) was over-expressed as a surrogate, because the true aldolase gene in *S. cattleya* could not be identified.

Full genome sequencing has however revealed four candidate aldolases: SCATT-10350 (216 amino acids), SCATT_p05770 (230 amino acids), SCATT-p01080 (208 amino acids), SCATT-33160 (340 amino acids) (see Supporting information S1). Of these genes, SCATT_p05770 is encoded by the *S. cattleya* mega plasmid shows a high homology (35% amino acid identity) to the original surrogate aldolase from *S. coelicolor* (SCO1844). To verify its putative role this gene was inactivated. PCR and subsequent sequencing was able to confirm the deletion (see Supporting information S2). However the mutant was still able to elaborate both of the fluorometabolites when incubated with fluoride ion (2 mM). This experiment suggests that aldolase encoded by SCATT_p05770 is not the aldolase involved in the conversion of 5-DRulP to fluoroacetaldehyde.

Another aldolase was purified *de novo* from *S. cattleya*. A cell free extract was subjected to two sequential ammonium sulphate precipitations. Purification by hydrophobic Phenyl-HP chromatography followed by size exclusion and then anion exchange chromatography resulted in a homogeneous protein. Although this aldolase could catalyse a reaction between fluoroacetaldehyde

and dihydroxyacetone phosphate, $^{19}F$ NMR analysis against reference compounds indicated that the catalysis generated the 5-fluoro-5-deoxyfructose diastereoisomer rather than 5-fluoro-5-deoxyribulose, and thus mediated a reaction by a different stereochemical course to that required for fluorometabolite production. Trypsin digestion and then Edman degradation analysis gave some amino acid sequence; "Phe-Ala-Tyr-Pro-Ala-Ile-Asn-Val-Thr-Ser-Gln-Thr-Leu-His-Ala-Ala-Leu-Arg". Genome scanning allowed identification of this aldolase gene too (SCATT_33160), among the four putative aldolases identified in the genome (see Supporting information S1). The *N*-terminal sequence is now reported as Met-Leu-Asp-Arg-Ala-Lys-Ala-Gly-Arg-Phe-Ala-Tyr-Pro-Ala-Ile-Asn-Val-Thr-Ser-Gln-Thr-Leu-His-Ala-Ala-Leu-Arg-Gly-Leu-Ala-Glu-Ala-Glu-Ser. The underlined residues match the Edman degredation sequencing and suggest that gene SCATT_33160 is a fructose aldolase. One of the two remaining aldolases (SCATT-10350 and SCATT-p01080) is likely to be involved in the fluorometabolite pathway.

*2.3.3. Inactivation of PLP-transaldolase (SCATT_p11780)*

The PLP-transaldolase which encodes a 635 amino acids protein in *S. cattleya* catalyses the reaction of fluoroacetaldehyde and L-threonine to generate 4-FT and acetaldehyde. The gene was previously identified by reverse genetics and the enzyme has been expressed [13]. The role of pyridoxal phosphate in mediating an exchange of L-threonine and fluoroacetaldehyde with 4-FT and acetaldehyde was confirmed with isotope labelled substrates [22]. Genome sequencing locates the PLP-transaldolase gene (*SCATT_p11780*) on the megaplasmid, and a gene knock out mutant study was carried out to further explore its role on the pathway. Firstly, the PLP-transaldolase gene was used as a DNA probe, and a fosmid (15D8) was identified from a *S. cattleya* genomic library, which harboured the transaldolase gene. Then, using the REDIRECT system as described in Section 4, an apramycin resistance and chloramphenicol sensitive strain of *S. cattleya* was constructed. PCR and subsequent sequencing confirmed the correct insertion (see Supporting information S2). The mutant was cultured in a media containing fluoride ion (2 mM). $^{19}F$ NMR analysis showed that this mutant had lost its ability to synthesize 4-FT but had retained the capacity to produce FAc at a comparable level to the wild type strain (Fig. 3). This outcome is entirely consistent with the PLP-transaldolase operating as the catalyst for the last step in 4-FT biosynthesis and that it is operating beyond the fluoroacetaldehyde branch point on the FAc/4FT pathway.

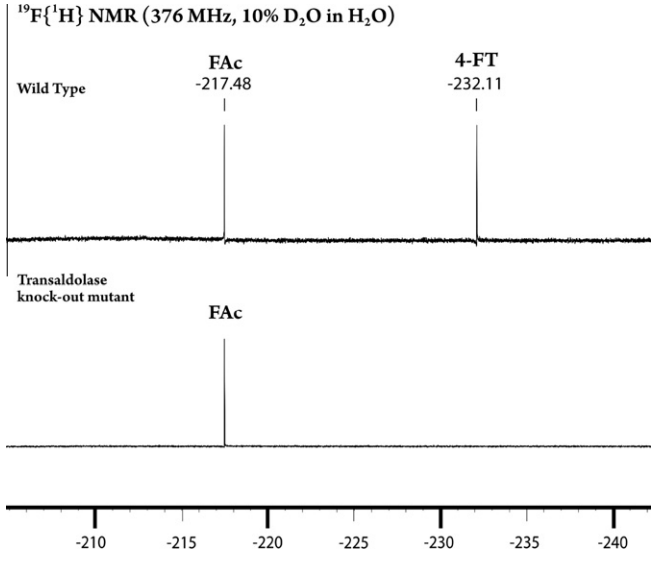



**Fig. 3.** $^{19}F\{^{1}H\}$ NMR spectra of wild type production of the fluorometabolites FAc and 4-FT, and the 4-FT-acetaldehyde transaldolase (*SCATT_p11780*) mutant which only produced fluoroacetate.

A cell free study with cells of *S. cattleya* was carried out to establish if any intermediate metabolites accumulated in this mutant strain. Accordingly the strain was grown without fluoride ion and the mycelium collected after a 7 d incubation. After ultrasonic treatment, the supernatant was incubated with SAM and fluoride ion and the organic fluorine products were analysed by $^{19}F$ NMR after a 20 h incubation. Both 5-FDRP and FDRulP as well as FAc were detected by $^{19}F$ NMR in the cell free extract, indicating that the mutant organism was otherwise competent in fluorometabolite biosynthesis except for its disabled transaldolase activity.

## 3. Conclusions

An original identification of the fluorinase gene located a 12 kb segment of genes (Spencer cluster) containing the fluorinase (*flA*) and the second biosynthetic gene (*flB*) [16]. The surrounding genes had no obvious candidates for the remaining biosynthesis genes, thus it appeared that they were scattered on the genome and not clustered like many other secondary metabolite pathways.

This study has benefited from full genome sequencing studies [21] of *S. cattleya*, and set out to locate the biosynthetic genes and confirm the role of key enzymes by knockout mutation. The first two genes *flA* and *flB* were knocked out and both mutants lost their capacity to biosynthesise the fluorometabolites as expected. Notably addition of 5′-FDA to the medium restored the ability of the *flA*$^{-}$ mutant to produce FAc and 4-FT. The other biosynthetic genes involved in 4-FT biosynthesis had to be identified after full genome sequencing, and in turn the 5-FDRP-isomerase and 4-FT acetaldehyde transaldolase were located. Most notably knockout of the transaldolase gene led to a mutant that lost the ability to 4-FT, but retained the capacity to biosynthesise FAc. This is entirely consistent with the transaldolase operating at a branch point of metabolism where fluoroacetaldehyde can be processed in two directions. It is perhaps surprising that such a key gene in 4-FT biosynthesis is located on the mega plasmid and significantly remote from the fluorinase (*flA*) gene. Genome sequencing revealed four aldolases. One aldolase (SCATT_33160) was established as a fructose aldolase after purification and assay. Edman degradation sequencing after trypsin digest gave sufficient amino acid sequence to locate this aldolase in the genome. Knockout of the aldolase (SCATT_p05770) with the highest homology to that in *S. coelicolor* did not perturb the biosynthesis of FAc and 4-FT. Thus the aldolase gene involved in the biosynthesis remains to be identified.

The *flH* gene, a putative $Na^{+}/H^{+}$ antiporter, located close to the fluorinase (*flA*) gene, was anticipated to play a role in fluoride ion uptake. However a *flH* knockout mutant, resulted in a healthy organism which was still able to produce FAc and 4-FT, indicating that this gene is not obviously required for fluorometabolite production. Knockout of the *flK* gene, the putative resistance gene to FAc toxicity, resulted in shutting down the fluorometabolite biosynthetic pathway. This gene expresses a fluoroacetyl-CoA hydrolase and perhaps the organism responds to its inactivation, by triggering an arrest of the pathway to avoid fluoroacetate toxicity.

In conclusion it is a striking feature of fluorometabolite biosynthesis that the genes are not clustered on the chromosome and thus regulation of this pathway would appear to be more complex than those of the clustered genes which are a feature of secondary metabolites in Actinomycetes. Some other secondary metabolite pathways in microorganisms are only partially clustered. For example ansamitocin [23], moenomycin [24] and meilingmycin [25] fall into this category. The fluorometabolite pathway joins this

relatively rare grouping of secondary metabolites where the genes are spread through the genome.

## 4. Materials and methods

### 4.1. Bacterial strains, culture conditions, and cloning vectors

*S. cattleya* DSM 46488 was used in this study. The *S. cattleya flA*, *flB*, *flH*, *flK*, isomerase, transaldolase, and the aldolase mutants were generated in this work. *E. coli* DH10B (Invitrogen) was used as a cloning host. *E. coli* ET12567 harbouring pUZ8002 was used as the donor for intergeneric bi-parental conjugation between *E. coli* and *S. cattleya* [26]. For sporulation, *S. cattleya* were grown on MS medium [26], and for antibiotic production, *S. cattleya* was cultured in a complex medium [4] (per litre): $KH_2PO_4$ (2.0 g), $NH_4C1$ (1.5 g), $MgSO_4{\cdot}7H_2O$ (0.5 g), NaCl (0.5 g), glycerol (10 g), inositol (0.4 g), monosodium L-glutamate monohydrate (5.0 g), NaF (0.084 g), $FeSO_4{\cdot}7H_2O$ (0.025 g), $ZnSO_4{\cdot}7H_2O$ (0.01 g), $CoCl_2{\cdot}6H_2O$ (0.01 g), $CaCO_3$(0.25), p-aminobenzoic acid (0.0001 g). The pH of the medium was adjusted to 7.0. Cosmid 8E11 and fosmid 15D8, 29C1 were used as a source of *S. cattleya* genes for inactivation. The vectors pIJ2925 and pOJ446 have been described in the literature [26]. When appropriate, antibiotics (thiostrepton, apramycin, kanamycin and chloramphenicol) were supplemented to a final concentration of 25 μg/mL for resistance selection. Nalidixic acid (50 μg/mL) was used for conjugation and 100 μg/mL ampicillin, kanamycin were used for *E. coli* selection.

### 4.2. DNA manipulations, sequencing, and analysis

Total DNA isolation, plasmid, DNA preparations, restriction endonuclease digestions, ligations, and other DNA manipulations were performed according to standard procedures for *E. coli* and *Streptomyces* [26]. Intergeneric conjugation from *E. coli* ET12567/pUZ8002 into *S. cattleya* was performed according to procedures provided by the John Innes Centre [27]. The sequencing service at Dundee University was used for DNA sequencing. Computer-assisted database searching and sequence analyses were carried out online using the frame-plot sequence analysis software and the BLAST program.

### 4.3. Genome sequencing, assembly and annotation

Whole genome sequencing of *S. cattleya* DSM 46488 was performed using 454 sequencing with the additional Solexa mate pair data. First, 581,262 reads with an average length of 414 bp were generated by using Roche 454 GS-FLX System (about 30-fold coverage). This resulted in 363 contigs. Then, a genomic library with 3 kb inserts was constructed and a total of 1,805,657 paired reads with an average length of 100 bp were obtained using the Illumina GAIIx Solexa System (about 45-fold coverage), leading to 11 large scaffolds. Finally, gaps were closed by sequencing PCR products with ABI 3730. Sequences were assembled with Phrap [28] and Consed [29]. The putative protein-coding regions were identified by Glimmer 3.0 [30] and assigned to different families according to the COG database [31]. The tRNA and rRNA genes were predicted by tRNAscan-SE [32] and RNAmmer [33], respectively. The chromosome and mega-plasmid sequences of *S. cattleya* DSM 46488 are available at GenBank under accession numbers CP003219 and CP003229, respectively.

### 4.4. Gene replacement experiments

Gene inactivations, except for the aldolase, were carried out by REDIRECT protocols provided by the John Innes Institute. In all cases, the *aac(3)IV* gene from pIJ773 was inserted within the coding region of the gene to be disrupted and in the same direction of transcription. The introduction of DNA into *S. cattleya* was achieved through intergeneric conjugation from *E. coli* ET12567/pUZ8002. After the introduction of the constructs into *S. cattleya* by intergeneric conjugation, exconjugants, in which a double crossover occurred, were identified by their resistance to apramycin and susceptibility to kanamycin (for cosmid 8E11) or chloramphenicol (for fosmid 29C1 and 15D8).

To generate a mutant containing a deletion or an insertion mutation, the following primers were designed: (i) Mutagenesis of *flA* (fluorinase): *flA*-FP: CGA GCG CGC CGA GGG CTC GTA CAT CTA CAT CGC GCC CAA att ccg ggg atc cgt cga cc, *flA*-RP: TAG CCG TGC TCC TCC AGC ACG GTG GTC AGC AGC CCG TTG tgt agg ctg gag ctg ctt c. (ii) Mutagenesis of *flB* (PNP): *flB*-FP: TCA GCC GCG CCG GGC GAA CCC CTT GCG CAG CAG CGC CGG att ccg ggg atc cgt cga cc, *flB*-RP: ATG CGG GCA CGG AAA TCG GGG AAC GAG CAG CGG AAC GCC tgt agg ctg gag ctg ctt c. (iii) Mutagenesis of *flK* (thioesterase): *flK*-FP: CGG TGC CGA GGC TGC CCT CGC CGG GTT CGA GGT AGG GCG att ccg ggg atc cgt cga cc, *flK*-RP: TGG TGG GGC TGA TGG AGT GGG CGT GCG TGC GGG CCA TGG tgt agg ctg gag ctg ctt c. (iv) Mutagenesis of *flH*: *flH*-FP: TCA TGC GGT TTC CGC CTT TCG CCT GGT GGT CTC GGG GTC tgt agg ctg gag ctg ctt c, *flH*-RP: GTG TCC CTG ACA ACC GTC GAA CTC GCC CAT GTC CTC ATC att ccg ggg atc cgt cga cc. (iv) Mutagenesis of isomerase: isomerase-FP: ATG GGT GAT CAG TCC GTA CAG CCT TTG GCC AAG GGC ACG att ccg ggg atc cgt cga cc, Isomerase-RP: TCA CGG CTG GGC GCG GAC GGG GCC GGG GCG GGC GGC CAG tgt agg ctg gag ctg ctt c. (v) Mutagenesis of PLP-transaldolase: transaldolase-FP: ATG CCG TCG TCC GTG AAC CGC ACC AGC CGC ACC GAG CCC att ccg ggg atc cgt cga cc, tranaldolase-RP: TCA GCC GGT GAT CCG GCG GAC GTC CTC GAT GAG GGC GAG tgt agg ctg gag ctg ctt c.

The apramycin resistance cassette pIJ773 was used as a template throughout [26]. The PCR products were transferred into competent cells of BW25113 (pIJ790, cosmid 8E11) for *flA, flB, flH, flK* or BW25113 (pIJ790, fosmid 29C1) for isomerase or BW25113 (pIJ790, fosmid 15D8) for PLP-transaldolase to isolate apramycin-resistant and amplicillin-resistant *E. coli* colonies for cosmid 8E11 or apramycin and chlorophenicol resistance *E. coli* colonies for fosmids. These mutated cosmids (fosmids) were introduced into ET12567 (pUZ8002) and then transferred by conjugation to *S. cattleya* DSM46488 to screen for apramycin resistant and kanamycin-sensitive colonies. This resulted in Δ*flA*, Δ*flB*, Δ*flH*, Δ*flK* mutants. Or for the Δisomerase and ΔPLP-transaldolase mutants apramycin resistant and chloramphenicol sensitive colonies were selected respectively.

### 4.5. Inactivation of the aldolase gene

Aldolase gene inactivation was carried out as follows. Two DNA fragments at each end of the aldolase genes were amplified by PCR from *S. cattleya* chromosomal DNA using the following primers; aldolase-1 (TGGTGCGGAGCGAACAGG), aldolase-2 (CCGTGAAGGC AGCCGTGAT), aldolase-3 (GAAGACACCGTGGCTGCG), and aldolase-4 (CATGGGCACCTCGACCTG). The 1.7 kb PCR product (from aldolase-1 and aldolase 2) and 1.46 kb PCR product (from aldolase-3 and aldolase 4) were cloned into a PCR-BLUNT vector (Novagen) followed by digestion with *Kpn*I + *Xba*I and *Xba*I + *Hin*dIII respectively and together they were ligated into *Kpn*I + *Hin*dIII site of vector pIJ2925.

The apramycin resistance gene from pIJ773 was inserted into the above construct (5.8 kb) *via* a PCR-targeting strategy with the following primers using the same method as described above. Aldolase-FP: CGG TGC GGG CCA CGT CCT CGC ACA TCA CGG CTG CCT TCA TGT AGG CTG GAG CTG CTT C, aldolase-RP: TCA CCG

TTC GCC CGC GGT GCT CAT GCG CAG CCA CGG TGT GCT CAC GGT AAC TGA TGC C. The ampicillin resistance and apramycin resistance plasmid was then digested with *Bgl*II and then ligated into the *Bam*HI site of the *E. coli–Streptomyces* shuttle vector pYH7 [34] to give a final construct for aldolase inactivation.

The above plasmid was then transferred into *S. cattleya* by conjugation and apramycin resistance and thiostrepton sensitive colonies were selected and confirmed as aldolase mutants by PCR and subsequent DNA sequencing.

### 4.6. Analysis of fluorometabolite production by $^{19}F$ NMR [4]

*S. cattleya* DSM 46488 and the mutants were grown in 500-mL flasks in a medium as previously described. After a 6 d incubation, the cells were harvested and washed with Tris buffer (50 mm, pH 7.8) and then resuspended in the same buffer (0.1 g wet-cell weight mL/1). Ultrasonication and centrifugation gave a supernatant that was used directly as the cell free extract. Biotransformations were initiated by supplementation of SAM (0.4 mM) and NaF (10 mM) to the cell-free extract (3 mL), and the reactions were incubated at 28 °C for up to 20 h [10]. For real-time analysis by $^{19}F$ NMR spectroscopy, samples were prepared in the same manner but $D_2O$ was added (100 μL) as an internal lock. $^{19}F$ NMR spectra were recorded on a Varian Inova 400-MHz NMR spectrometer. FAc and 4-FT production was determined by reference to established chemical shifts in $^{19}F$ NMR.

## Acknowledgments

We greatfully acknowledge grants from the BBSRC (BB/F007426/1) and grants from the 973 Program (2009CB118901, 2012CB721002), Ministry of Science and Technology, China; the National Natural Science Foundation of China; the Program for New Century Excellent Talents in University, Ministry of Education, China (NCET-10-0572); the Chen Xing Young Scholars Program.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bioorg.2012.06.002.

## References

[1] H. Deng, D. O'Hagan, C. Schaffrath, Nat. Prod. Rep. 21 (2004) 773–784.
[2] (a) W. K Hagmann, J. Med. Chem. 51 (2008) 4360–4369;
(b) I. Ojima, Fluorine in Medicinal Chemistry and Chemical Biology, J. Wiley & Son, Chichester, 2009;
(c) A. Tressaud, G. Haufe, Fluorine in Health: Molecular Imaging, Biomedical Materials and Pharmaceuticals, Elsevier, Amsterdam, 2008.
[3] (a) F.H. Vaillancourt, E. Yeh, D.A. Vosburg, S. Garneau-Tsodikova, C.T. Walsh, Chem. Rev. 106 (2006) 3364–3378;
(b) A. Butler, M. Sandy, Nature 460 (2009) 848–854;
(c) C.D. Murphy, Nat. Prod. Rep. 23 (2006) 147–152.
[4] (a) M. Sanada, T. Miyano, S. Iwadare, J.M. Williamson, B.H. Arison, J.L. Smith, A.W. Douglas, J.M. Liesch, E. Inamine, J. Antibiot. 39 (1986) 259–265;
(b) K.A. Reid, J.T. Hamilton, R.D. Bowden, D. O'Hagan, L. Dasaradhi, M.R. Amin, D.B. Harper, Microbiology 141 (1995) 1385–1393.
[5] D. O'Hagan, C. Schaffrath, S.L. Cobb, J.T. Hamilton, C.D. Murphy, Nature 416 (2002) 279.
[6] C. Dong, F. Huang, H. Deng, C. Schaffrath, J.B. Spencer, D. O'Hagan, J.H. Naismith, Nature 427 (2004) 561–565.
[7] M. Rodriguez, L.E. Nunez, A.F. Brana, C. Mendez, J.A. Salas, G. Blanco, Antimicrob. Agents Chemother. 55 (2011) 1638–1649.
[8] M.J. Bodner, R. Li, R.M. Phelan, M.F. Freeman, K.A. Moshos, E.P. Lloyd, C.A. Townsend, ChemBioChem 12 (2011) 2159–2165.
[9] S.L. Cobb, H. Deng, J.T. Hamilton, R.P. McGlinchey, D. O'Hagan, Chem. Commun. (2004) 592–593.
[10] M. Onega, R.P. McGlinchey, H. Deng, J.T. Hamilton, D. O'Hagan, Bioorg. Chem. 35 (2007) 375–385.
[11] C.D. Murphy, S.J. Moss, D. O'Hagan, Appl. Environ. Microbiol. 67 (2001) 4919–4921.
[12] C. Schaffrath, C.D. Murphy, D. O'Hagan, Angew. Chem. Int. Ed. 40 (2001) 4479–4481.
[13] H. Deng, S.M. Cross, R.P. McGlinchey, J.T. Hamilton, D. O'Hagan, Chem. Biol. 15 (2008) 1268–1276.
[14] S.L. Cobb, H. Deng, J.T. Hamilton, R.P. McGlinchey, D. O'Hagan, C. Schaffrath, Bioorg. Chem. 33 (2005) 393–401.
[15] C. Schaffrath, S.L. Cobb, D. O'Hagan, Angew. Chem. Int. Ed. Engl. 41 (2002) 3913–3915.
[16] F. Huang, S.F. Haydock, D. Spiteller, T. Mironenko, T.L. Li, D. O'Hagan, P.F. Leadlay, J.B. Spencer, Chem. Biol. 13 (2006) 475–484.
[17] (a) A.S. Eustaquio, F. Pojer, J.P. Noel, B.S. Moore, Nat. Chem. Biol. 4 (2008) 69–74;
(b) A.S. Eustaquio, B.S. Moore, Angew. Chem. Int. Ed. Engl. 47 (2008) 3936–3938.
[18] A.S. Eustáquio, D. O'Hagan, B.S. Moore, J. Nat. Prod. 73 (2010) 378–382.
[19] (a) M.V. Dias, F. Huang, D.Y. Chirgadze, M. Tosin, D. Spiteller, E.F. Dry, P.F. Leadlay, J.B. Spencer, T.L. Blundell, J. Biol. Chem. 285 (2010) 22495–22504;
(b) A.M. Weeks, S.M. Coyle, M. Jinek, J.A. Doudna, M.C. Chang, Biochemistry 49 (2010) 9269–9279.
[20] E. Kirsten, M.L. Sharma, E. Kun, Mol. Pharmacol. 14 (1978) 172–184.
[21] V. Barbe, M. Bouzon, S. Mangenot, B. Badet, J. Poulain, B. Segurens, D. Vallenet, P. Marliere, J. Weissenbach, J. Bacteriol. 193 (2011) 5055–5056.
[22] C.D. Murphy, D. O'Hagan, C. Schaffrath, Angew. Chem. Int. Ed. Engl. 40 (2001) 4479–4481.
[23] T.W. Yu, L. Bai, D. Clade, D. Hoffmann, S. Toelzer, K.Q. Trinh, J. Xu, S.J. Moss, E. Leistner, H.G. Floss, Proc. Natl. Acad. Sci. 99 (2002) 7968–7973.
[24] B. Ostash, A. Saghatelian, S. Walker, Chem. Biol. 14 (2007) 257–267.
[25] Y. He, Y. Sun, T. Liu, X. Zhou, L. Bai, Z. Deng, Appl. Environ. Microbiol. 76 (2010) 3283–3292.
[26] T. Kieser, M. Bibb, J.M. Buttner, J.K.F. Chater, D.A. Hopwood, Pract. Streptomyces Genet. (2000).
[27] B. Gust, G.L. Challis, K. Fowler, T. Kieser, K.F. Chater, Proc. Natl. Acad. Sci.100 (2003) 1541–1546.
[28] B. Ewing, L.D. Hillier, M.C. Wendl, P. Green, Genome Res. 8 (1998) 175.
[29] D. Gordon, C. Abajian, P. Green, Genome Res. 8 (1998) 195.
[30] A.L. Delcher, K.A. Bratke, E.C. Powers, S.L. Salzberg, Bioinformatics 23 (2007) 673.
[31] R. Tatusov, N. Fedorova, J. Jackson, A. Jacobs, B. Kiryutin, E. Koonin, D. Krylov, R. Mazumder, S. Mekhedov, A. Nikolskaya, BMC Bioinform. 4 (2003) 41.
[32] P. Schattner, A.N. Brooks, T.M. Lowe, Nucl. Acids Res. 33 (2005) W686.
[33] K. Lagesen, P. Hallin, E.A. R dland, H.H. St. rfeldt, T. Rognes, D.W. Ussery, Nucl. Acids Res. 35 (2007) 3100.
[34] Y. Sun, X. He, J. Liang, X. Zhou, Z. Deng, Appl. Microbiol. Biotechnol. 82 (2009) 303–310.