This journal is © The Royal Society of Chemistry 2007

# Characterisation of 3-Methylorcinaldehyde Synthase (MOS) in *Acremonium strictum*: First Observation of a Reductive Release Mechanism During Polyketide Biosynthesis.

**Andrew M. Bailey,[b] Russell J. Cox,*[a] Kate Harley,[a] Colin M. Lazarus,[b] Thomas J. Simpson*[a] and Elizabeth Skellam.[a]**

**Electronic Supplementary Information**

## 1. Experimental Procedures.

### 1.1 PCR Primer Sequences.

KHKS2 and KHKS3c degenerate oligonucleotide PCR primers were designed using multiple alignments of non-reducing clade III KS domain sequences obtained from public databases (see section 3 below).

```
KHKS2:    5'-GCIGAYGGITAYTGYMGIGG-3'
KHKS3c:   5'-GTICCIGTICCRTGIGCYTC-3'

Y = T or C; M = A or C; R = A or G; I = inosine.
```

### 1.2 gDNA preparation and library construction.

*Acremonium strictum* gDNA was prepared using a method for producing library quality DNA which avoids the use of a caesium chloride density gradient. The resulting gDNA was used to construct a genomic library. The commercially available Lambda phage vector system (λBlueSTAR; Novagen) was chosen for this purpose. Thus, the prepared genomic DNA was partially digested using *Sau*3AI and ligated into *Xho*I arms supplied with the λBluestar kit. The phage library was packaged using the Ready-to-Go Lambda Packaging Kit (Amersham Biosciences) and assayed by plating with KW251 *E. coli* cells. Once successful small-scale ligation and packaging conditions had been achieved, the reactions were scaled up to yield ~10,000 phage clones which were plated out (23 x 23 cm plates). The plaques were visible after incubation (37°C, 16 hours) and transferred to two sequential Zeta-Probe® GT membranes (BioRad).

### 1.3 PCR Conditions, Library-Probing and Sequencing.

Oligonucleotide-specified sections of DNA were amplified by PCR using Thermoprime Plus DNA polymerase (ABgene®), giving 3´A overhung products such that PCR products can be ligated directly into pGEM-T Easy (Promega) or pCR®2.1-TOPO® (Invitrogen) cloning vectors, in a final volume of 25 µl. 2 x Reddy Mix™ PCR Master Mix (ABgene®) which contains Thermoprime Plus DNA polymerase (0.625 units), Tris-HCl pH 8.8 (75 mM), $(NH_4)_2SO_4$ (20 mM), $MgCl_2$ (1.5 mM), Tween 20 (0.01% v/v), dATP, dGTP, dTTP, dCTP (0.2 mM each) and precipitant plus red dye for electrophoresis. Primers at 0.1-1.0 µM, additional $MgCl_2$ (0-5 mM; Sigma) and 10-100 ng genomic DNA template were also added.

Samples were centrifuged and amplified in a Programmable Thermal Controller (Hybaid Ltd, PCRSprint). The following thermal cycling profile was run: Initial denaturation at 94 ºC for 3 minutes; 10 cycles of denaturation at 94 ºC (15 seconds), primer annealing at 45-55 ºC (30 seconds), extension at 72 ºC (1 minute per kb of DNA to be amplified - usually 45 seconds); Denaturation at 94 ºC (15 seconds), primer annealing at 45-55 ºC (30 seconds), extension at 72 ºC (48 seconds for the first round

and an additional 3 seconds for every additional round, for a total of 20 rounds; Final extension at 72 ºC (6 minutes); Cooling to 4 ºC.

The PCR product was radio-labelled and used to probe the genomic DNA library. Two independent clones were isolated from the *Acremonium strictum* genomic DNA library (KHIII4A2 and KHIII4B1) and these were sequenced outwards in both directions from the region relating to the location of the PCR product obtained with our degenerate primers (Lark Technologies). Sequencing was continued until entire PKS had been sequenced. The results of the sequencing reactions were aligned to form a contig assembly which contained at least a 500 bp overlap between fragments.

The sequence has been deposited with the EMBL database, Accession#: AM745350.

## 1.31 DNA Sequence of *ASpks1*

```
   1     ATGGCAGCTC ATGGGCAAAC CTCAAAACGG GGTAACAACA CCCTGCTGCT CTTCGGGGCC CTCGTTCAAT CACACGATGT GTCCACGTTG AGGAGTATGC
 101     GTGAGTCCAT TGTAGTACAA CATGGGGAAC ACAGCTGGTT GGTGGACTCA ATCAAGGCCC TACCCCAGGA CTTTGAAGCA GCCCTCCCAC ATCTACCATT
 201     TTTCGACCAA GCAACAACAA CGACAATTCA TCAGTTGCTT GTCGATGCCG TGTCAAGTTT CTTGACCGGA TCTTTCGAAA CTCTCGTCTC GCCCTTGCCC
 301     GCTGCCCTGC TGATTCCCCT GGCCGTTGCC ACACAGCTCG CGCACTATGT CGAGTATTCA CGCCAGTCTC CAACTGGTTT GGCCGAGGGC AAAGAGGCAC
 401     TCGGGTTTTG CACAGGCATC TTGAGTGCCT TTGCTGTGGC CAGTTCTCAT GATGTCTGTG ATTTGGCCAA GTACGGTGCG GCTGCCATGC GTCTAGGTAT
 501     GCTCGTGGGT CTGGTCGTGG ACTGCGAGGA TGCAGCAGCT GGTCAGGGGA GATACCGATC TGTGAGCGCC GGCTGGGATT CGGAAGAGAA ACATGCTGCG
 601     ATGCTCAAGA TTGTTCAAAG CTTTGAAGAG GTGAGCATCA TATGTCCATG TCTACCTGAG TCAAAATGCA GACGACTAAT CGCATCATTT ATCTACAGGC
 701     ATACGTCTCG GTGCACTTTG ACAAGAACCG CGCCACAATC ACAACCAGCC CAGGGACCAT ATCCAACTTG ACCCGCCAAC TGCAAAAAGA AGGCCTCGTA
 801     GCCTCGGACA TGGGTCTCCT CGGCCGCTTC CACTTTGCCG GAAGCACCAA GCCCCGCGAA GTCACAGTGG ACCAGCTCGT TAGCTTTTGC AACTCGCCAG
 901     CTGGCGCACT GTTCCGACTC CCAGATGCAG ACTCACTCCG TCTTGCAACG CGTATCAATG ACAGGGATGG CGGGCTTATC ACTCAGGGTT CGTTACATGA
1001     ACACGCGCTG CAATCCATCC TGGTCAAACT GGCGGCATGG TTTGAGACCT TCTCATCCGC CACGACAACA CAAGCCAACA CGGGGGCACA GAATGGTCGA
1101     GCTCGACCTC AAATTGTCGA CTTTGGTCCC CAAAACAGCG TTCCGCACTC CCTGGCCTCT ACAGTTGACA TCAATTCAGG CAATGGCAAG ACTAGGCGAG
1201     TCAAGCCAGC AGACGCCCAG TCTTCTGCTA ACTCCACCCA TACCCGCCCC TGGCTCGACA CTGATATTGC CATAGTAGGC ATGTCCTGCA AGGTCCCTGG
1301     AGCCGAGAAC CTTGAGGAGT TCTGGGATTT GCTGGTCTCT GGCAAGTCAC AGCACCAAGA AATCTCAGGG CAAGAAGGGG GACGCTTCGA CTTCGGTGAC
1401     ACAGCTTTCC GCACTGCCGC CGACCAACGA CGTCGATGGT TCGCCAACCT GGTCTCAAAC CATGATCAGT TCGATCACCG CTTCTTCAAA AAGTCTGCGC
1501     GTGAGAGCGC GTCGATGGAC CCTCAACAGC GGCACATACT CCAGGTAGCG TACCAGGCTG TAGAAGGAAG CGGCTACTTC AACAAGTCCT CCTCCTCAAC
1601     CCCCACAAAC GCAAACATTG GGTGTTACGT GGGGCTCTGC CTAGGAGATT ACGAGTCCAA TGTGGCTTCG CATCCCGCGA CCGCTTTCAC AGCAACCGGA
1701     AACCTCCAAG GCTTCGTCTC AGGCAAAGTC TCGCACTACT TCGGGTGGAC GGGGCCAGCA GTCACGGTCA ACACAGCGTG CAGCTCTTCC CTGGTTGCCG
1801     TGCACCTAGC CTGCCAGGCT ATCTTGTCAG GGGAGTGCGA GGCTGCGCTG GCTGGAGGTT CGCACATCAT GACGTCCGCA ACGTGGTTTC AGAACCTCGC
1901     AGGAGGGTCG TTCCTGAGCC CTACAGGGGC CTGCAAGCCG TTCGACTCCA AAGCTGATGG GTACTGCCGC GGTGAGGGTG TGGGAGCTGT CTTCTTAAAG
2001     AGGATGAGCC AGGCCATGGC CGATGGTGAT ATGGTCTTGG GTGTGGTGGC CGCGACGGGT GTCCAGCAGA ACCAGAACTG CACCCCTATC TTCGTGCCCA
2101     ACGCACCTTC ATTGGAGAAC TTGTTTAGCC GCGTCATGAC CAAAGCACGT GTTAAGCCAG CAGACATCTC GGTTGTTGAG GGTCATGGCA CGGGTACGGC
2201     GGTGGGAGAC CCTGCTGAAT ACGATGCCAT TCGGAAAGCT CTTGGAGGCA CTACTCATCG CTCAGCCGAC AAGCCTCTTA TGCTCAGCTC AGTCAAAGGA
2301     CTTGTGGGCC ATATGGAGTG CACATCAGGG GTCATTGGTA TGATCAAACT CTTGTTGATG ATGAATAAGG GCGCCCTGCC ACCCCAGGCT AGCTTCCAAA
2401     GCATCAACCC GGCCCTTGGA GCGACGCCCG CCGACCACAT GTTTATACCT ACCCGGCCTC AGCCATGGGT GGTGCCGGCA GGCGGATTCC GAGCTGCTCT
2501     GCTCAACAAC TACGGTGCTT CGGGGTCAAA CGCTTCTGCA GTCCTTGTCC AGTCACCTTC AATGAGCTTC AGGCCCGAGA TCACAGTGGG GTCTAGGCCC
2601     GCAGCCGGGA TCAAGTTTCC TTTCTGGCTC GCGGCCTTTG ACAAGAAAAG CTTGAGCCGT TATGTGAAGG CTTTGCGCAA GTGGCTGTGC CGATTAGATG
2701     GAGATCAATC ACTCGCCAGC CTCTCGTTCA ATCTCGCACG CCAGAGCAAC CGCACAATGC AGGCCAACTT GGTCCTCACT GCCCGGTCTA TAGAAGCACT
2801     TGATCAGAGC CTGGCTGACT TTGAGAATGG AAATGATGGA TCTTTCATCG AAAGAACTCC AGCTTCCTCC CAGCCGACGG TGATCCTATG CTTTGGAGGT
2901     CAGGTTTCAT GTTTCGTGGG TCTTGACAAG CAAGTCTATC AAGACATGGC ACTTGTGCGC TATTATCTCG ACCGAGTTGA TGCTGTCATC CAGTGTCAGG
3001     GAGGCCGAAG TATCTTCCCA GGAATTTTCA ACCGCTCGCC TCCATCAAAG GTCGATATTG TGCATCTCCA TACGATGCTC TTCGCCATGC AATACGCCTC
3101     CGCACGCTGC TGGATCGACT CGGGTGTGAA GCCAGCCGCA CTTGTAGGCC ATAGTTTCGG AACCCTCACT GCCCTATGCA TCTCGGGCAT TCTTTCGCTT
3201     GAGGATACGA TCAAGGCCAT TATGTGCCGA GCCAAACTTC TCAACGAGGC CTGGGGCCCC GACCAAGGCG GCATGATTGC AGTCGAAGGG GATATCGACG
3301     TGATTGAAGA ACTCCTTGAT GAAGCGAACA AGAACCACGA CGACAAGCCC GCGACAATCG CCTGCTACAA TGGCCCGACG AGCTTCACAC TTGCTGGATC
3401     GACTACTGCT ATGGATGCCG TGGCTGCTCA GCTGAAAAAC GGCGCCAAGT ACAGCAAAGG CATGAAGTCT AAGAGGATCT ATGTCACACA CGCTTTCCAC
3501     TCGGTTCTTG TTGACCCTTT GCTCGAGGAG CTGACGCAAA GGGTGGCCGA TTCTGGAGTA AGGTTCCGCA AACCCATAAT TCCTGTCGAG CTGAGCACAG
3601     AACAACACAT GAGCGAAAGT GAGCTCACAA GTGAGTTTAT GGCAAACCAC ATGCGCCAGC CTGTGTACTT CCACCACGCC GTTGAGCGGC TCGCACGGAG
3701     GTATGCCGGA GGCTCCTCCC CATGTGTTTT CCTTGAGGCT GGTACGAACT CATCTGTCTG TAATATGGCA TCTCGTGCTC TAGGCAGCAC AGAGTTCGTC
3801     ACCAAGAGTT CCTCCCTCTC ATTCCATGGA GTCAACATCG CCAACTGCGA CGCCGGATGG AACAAGCTCA CCGACACCAC CGTAAATCTG TGGGAGACAG
3901     GTGTAAGAGT GCACCACTGG GCACATCACG GTGTACAGCA GATGCACCAG ACAGACATCA AGCCCTTGTT GGTACCTCCC TACCAGTTCG ACCCTGATTC
4001     GCGTCATTGG ATAGATCTCA AGGTTCCGCG CAAAGCTCTC ATGGAAACAG ATGAGGCAGA TGCGGGAGGC AAAAAGCAGT CGGATGCCGA AAAGCTACCA
4101     GAGACCATCT TGACCTTCCA CAGCTCTGAC GCTGTTGGGG CACAGAAGCA AGCACGATTC CGCGTCAACA CAATGCTAGA AGAATACAAG CAACTTCTGC
4201     GAGGACACAT GACTCTGGAG ACCGCACCCA TCCTGTCCGC CACCTTGCAG ATCAATCTCG TGATCGAGGC CATTTCCAGC ACTCAACCAG AGTACAAGTC
4301     ATCCAAGAGC CAGCCTCAGA TCCAAGATGT GGTCTACCAG TCACCTGTCT GCTTCAATTC GGCAAATACA TTATGGGTTG AGGTTACCAA CGTTAGTGGA
4401     CAATGGATGT TCCAAGTATT CAGCACTACC ACTCAGGAGC TTTCTCCCAA GTCGACACGT ATGGTACATA CCAAGGGCAC GGTTGCATTC AAGAACCCAG
4501     GGGATGCGGA AATCCGGCGT CAACTGATGA GCTACGAGCG TCTCTTCAGC CATGGCCGAG CGACAGACCT TCTACAGAAC AGCAATGCCT CGACTGCGCC
4601     CATTGACGAG ATGCTGGGCA ATCAAAGTAT CTACCGCATC TTCTCCGAGA TAGTAAGTTA TGGTCCCGAG TTCCGAGGCC TACAGAAGAT GGTGTCTCGG
4701     GGCAACGAAA CGGCAGGACA CGTCGTCCAT CTCAAGCATC AAGACTCGGC CAGTACCGAA GCCGAGCCAT GGTTCGACCC TCACTTGGCC GATACTTTCT
4801     GCCAGCTCGG TGGCTTGTGG GTCAACTGCA TGATGCCAGA GCGGGAGCGT GGAAATGGTC ATGTGTATCT GGCAAACGGA ATCGATCAGT GGATTCGCGG
4901     TTATCCCGCG GCCAGCACTG ACCGGCCCGA GGCTTTCAAC GTGTTTGCTG TGAATAAGCA GGCCTCAGAG CAGCTCACTT TGACTGATGT CTTCGTTTTC
5001     AATGCCGCGG ACGGGGCTCT TGTGGAAGTC ATCCTCGGTA TTGCGTATGT CAAGATCGCA AGGCCTTCAA TGGAGAAGTT GCTTGCCCGC TTGACAGAGC
5101     CTTCATGGGT GGCAGGTGGA AAGACTACGC CTCAGACTGC TACAAAACCA GCTGCAGCAC CTGTAGTTGC GGATCACACG CCTCGCACAA CAGAGTCAGC
5201     GTCAACAGTT AACGGGGTGA ACTTGGACGA TAGGAAGCCT GAAGGGACGG CACTCCCCCA AGAAATGCTC TCTGATACCG AAGAGCTCAG GCCCAAGGCA
5301     CAAGGACAAG AACTCCAAGA CATGATCGCC CGCGTCAAAG CCGTGATGGC TGACATTTCC GGCCTTGACA TTTCCGAAAT CAAGGACGAT AGCAACCTCG
5401     CTGACCTCGG CATCGACTCA CTCGTTGGTA TGGAAATGAC GCACGAGATT GAGTCGACTT TGAAGGTGGA ACTTCCCGAG TCCGAGATTA TGTCTGTTGT
5501     TGACATGGAA GGCCTGCTGC AATGCGTAGC TGGAGCATTG GGACTGTCGA TGACTGGAGC TTCAAGCGAC ACGTTAACAG CAAGCTCAGA CAGCGGCATT
5601     AATAGTGCCA AGTCATCCAT CCTGAGCGGT ACTTCGACGA GCACAAGCAC GGGTACTACA GATACGGGAT CTGATGTCGG CCAGAGTATG AAAGAGCCCT
5701     CTCTCATGTT GGACACAGTA AAGAAAGCCT TCGCACAGAC CAAAGAGGCA ACAGACGCCC GCATTAAAGC AGCCAGCAAC CAAGTTTCCT ACTGCTCTAC
5801     CTCCCTCCCG CAGCAGAATG AGCTATCAGT CTTGCTCACT ATAACAGCCC TTGAAGCCCT CGGTGCAGGT TTCAGCACTG CTCGTCCAGG CTCACAGCTT
5901     ACACGAATCT CCCACGCACC TGGACACGAG CAATTCGTGA CGCACCTGTA CAAAGAGATC GAGACAGCAA CCCAAATCAT CAAGATCGAC GGGCATGGCG
6001     CACAAGCTGT GATCACAAGG ACAGCTGTGC CGTTGCCAGA CGTGGAGAGC CGCCAGGTCG CGTTGTGCGA GCAGATGCTG AGAGGCGATC CCGAGCAAGT
6101     TGGTACAATG GAGTTGATCA AACATGCAGG CGAGAACCTT CACCGGGTGT TGAGCGGTGA GACAGATGGG GCTAAGGTCA TCTTCGGTAG CAAAACTGGC
6201     TCGAAGCTCG TTAGTCAGTG GTACGCGCAG TGGCCACTGA ACAGGTCATT GATTGCCCAG ATGGGAGATT TCCTTACTGC TGTCGTCGCC GGGATACAGG
6301     CTGACGAGGA CATGCCGTTC AGTGAAATTA ACCCCTTGCG TATCATGGAA ACAGGAGCCG GCACGGGGGG CACAACGAAA CAGATCGTGC CACTCCTAGC
6401     ACGCCTTGGT CTGCCCGTCG TGTACACTTT CACGGACCTT GCGCCATCTT TCGTGGCAGC GGCGCGCAAG ACCTGGGGTA AAGAGTACCC GTGGATGCAG
6501     TTCCGAACCC TCGACATGGA AAAGACGCCA CCATCCGTCG AGGATGGCCT GCCCCTGCAG CACTTCATCG TCTCAGCCAA CGCTGTTCAC GCGACCAAAT
6601     CGATCAGCGC GACGACAGGC AACCTACGCA AGGCGCTGCG CACAGACGGC TTCCTCCTGA TGATGGAGAT GACTCGGACA CCGTTCTGGG TAGACTTGAT
6701     TTTCGGACTT TTTGAAGGCT GGTGGTTGTT TGAGGACGGA AGGAAGCATG CGCTTACGCA TGAAGCGCTG TGGGATCAGG AGCTCAGCAA GGTTGGGTTT
6801     GGATATGTGG ATTGGACGGA GGGGATGACG GCTGAGAGTG AGATTCAGAA GATCATCTTG GCGTCAGCGG ACGCAAACAC TCGGTAATGG TTTTTCCTCC
6901     TTGTCTCCTA AGTATGTTGT GGACTCAATT CTAAGATCAT ATTTTTCTAC CATGACTAGC CTCGAACGAG TGCGCTTGCC AGCCAGCCAT ACCGACTACC
7001     ATTTGAATCA AGTTGGCGTG GAGAACGAAG CCCGAGAACT TATGGTTGCC GACTACGTCT CAACATTGAC CAAGGAGTTC AACAAGACCA TGACTCAGTA
7101     CACAGATGCG GGACTCTCCC TTTCCAGTCG CACATCTCAG ACACCTATGA GTAGCCAGAA AAGATGTATC CTCATCACGG GAGGCACCGG CGGCTTAGGT
7201     GCCCACCTGG TAGCGGAAGC CGCCCTCCTC CCCGACGTCA ATATGGTGAT ATGCCTGAAC CGACCAAACA GGAAACAAGA AGCACGAGAA AGACAGTTGG
7301     TATCCCTGGA AAAGAAAGGA CTCATTCTGA GTCCAGAGGC TCTGGCCAAG ATTACCGTGT TCGAAACGGA CCTTTCCCAG CCAGGTAGTC TCGGTCTCTC
7401     GGATGACAAG TATAACCTCC TGAGAGGTAA TGTGACGCAC ATCATCCACA ACGCCTGGCT GATGCACTCC AAATGGCCTG TCAGGCGTTT CGAACCGCAG
```

```
7501      CTACGCATAA TGGCGCACAT GTTGAACCTG GCAGCAGATA TTGCTACATG TCAACGCACC CAGGGGCAAC GTCAGCCAGG ACCACCGGTC TCCTTTGTCT
7601      TCGTCTCCTC TATCGCCACG GTAGGCTATC ATCCGGTCGT CACCAATCCA GGCAATCCTG CTGTCCCAGA GACACGGATC CCCATATCTA GCGTCTTACC
7701      GACAGGATAC GGTGAGGCAA AGTACATCTG CGAGCGTATG CTTGACGCGA CGCTTCATCA GTACCCGGCG CAGTTCAGAG CATCGGCGGT GCGACTGGGA
7801      CAGATCGCAG GCAGCGAGAT CAACGGCCAT TGGAACTCTG CGGAGCACAT CTCGTTCCTC GTTAAATCAT CACAGAGCAT CGGAGCCTTG CCTGCACTCC
7901      CGGGTCCAAT GGGCTGGACG CCAGCCGACT ACGTGGCTCG TGGGCTGGTA GAGATTGCGA CGCAACCGGA CAACATCGAA CTCTACCCAA TCTATCACAT
8001      TGAGAACCCC GTCAGGCAAC CCTGGGACGA GGCGCTGGCC GTTCTAGCCG ACGAGATGGG GATATCCTCA GAGGCGCTTC CGTTCCAGGA GTGGGTGCAA
8101      ACAGTGCGGG ACTGGCCACG ACAAGGGGAC AACACCGCCG CGGGCGCCAA CCCGGCGTAT CTCCTCGTGG ATTTCCTCGA GGATCACTTT CTACGGATGA
8201      GTTGTGGAGG TTTGCTACTG GGCACGGCGA AGGCGAGGGA GCATTCCCCG AGTCTTGCTG GGATGGGCCC GGTGAGCGAT GAGTTGCTTA GATTGTTTGT
8301      TAGAAGCTGG AAGGAGGTGG GGTTCTTGTT GTGA
```

## 1.32 Peptide Sequence of MOS.

```
   1 MAAHGQTSKR GNNTLLLFGA LVQSHDVSTL RSMRESIVVQ HGEHSWLVDS
  51 IKALPQDFEA ALPHLPFFDQ ATTTTIHQLL VDAVSSFLTG SFETLVSPLP
 101 AALLIPLAVA TQLAHYVEYS RQSPTGLAEG KEALGFCTGI LSAFAVASSH
 151 DVCDLAKYGA AAMRLGMLVG LVVDCEDAAA GQGRYRSVSA GWDSEEKHAA
 201 MLKIVQSFEE AYVSVHFDKN RATITTSPGT ISNLTRQLQK EGLVASDMGL
 251 LGRFHFAGST KPREVTVDQL VSFCNSPAGA LFRLPDADSL RLATRINDRD
 301 GGLITQGSLH EHALQSILVK LAAWFETFSS ATTTQANTGA QNGRARPQIV
 351 DFGPQNSVPH SLASTVDINS GNGKTRRVKP ADAQSSANST HTRPWLDTDI
 401 AIVGMSCKVP GAENLEEFWD LLVSGKSQHQ EISGQEGGRF DFGDTAFRTA
 451 ADQRRRWFAN LVSNHDQFDH RFFKKSARES ASMDPQQRHI LQVAYQAVEG
 501 SGYFNKSSSS TPTNANIGCY VGLCLGDYES NVASHPATAF TATGNLQGFV
 551 SGKVSHYFGW TGPAVTVNTA CSSSLVAVHL ACQAILSGEC EAALAGGSHI
 601 MTSATWFQNL AGGSFLSPTG ACKPFDSKAD GYCRGEGVGA VFLKRMSQAM
 651 ADGDMVLGVV AATGVQQNQN CTPIFVPNAP SLENLFSRVM TKARVKPADI
 701 SVVEGHGTGT AVGDPAEYDA IRKALGGTTH RSADKPLMLS SVKGLVGHME
 751 CTSGVIGMIK LLLMMNKGAL PPQASFQSIN PALGATPADH MFIPTRPQPW
 801 VVPAGGFRAA LLNNYGASGS NASAVLVQSP SMSFRPEITV GSRPAAGIKF
 851 PFWLAAFDKK SLSRYVKALR KWLCRLDGDQ SLASLSFNLA RQSNRTMQAN
 901 LVLTARSIEA LDQSLADFEN GNDGSFIERT PASSQPTVIL CFGGQVSCFV
 951 GLDKQVYQDM ALVRYYLDRV DAVIQCQGGR SIFPGIFNRS PPSKVDIVHL
1001 HTMLFAMQYA SARCWIDSGV KPAALVGHSF GTLTALCISG ILSLEDTIKA
1051 IMCRAKLLNE AWGPDQGGMI AVEGDIDVIE ELLDEANKNH DDKPATIACY
1101 NGPTSFTLAG STTAMDAVAA QLKNGAKYSK GMKSKRIYVT HAFHSVLVDP
1151 LLEELTQRVA DSGVRFRKPI IPVELSTEQH MSESELTSEF MANHMRQPVY
1201 FHHAVERLAR RYAGGSSPCV FLEAGTNSSV CNMASRALGS TEFVTKSSSL
1251 SFHGVNIANC DAGWNKLTDT TVNLWETGVR VHHWAHHGVQ QMHQTDIKPL
1301 LVPPYQFDPD SRHWIDLKVP RKALMETDEA DAGGKKQSDA EKLPETILTF
1351 HSSDAVGAQK QARFRVNTML EEYKQLLRGH MTLETAPILS ATLQINLVIE
1401 AISSTQPEYK SSKSQPQIQD VVYQSPVCFN SANTLWVEVT NVSGQWMFQV
1451 FSTTTQELSP KSTRMVHTKG TVAFKNPGDA EIRRQLMSYE RLFSHGRATD
1501 LLQNSNASTA PIDEMLGNQS IYRIFSEIVS YGPEFRGLQK MVSRGNETAG
1551 HVVHLKHQDS ASTEAEPWFD PHLADTFCQL GGLWVNCMMP ERERGNGHVY
1601 LANGIDQWIR GYPAASTDRP EAFNVFAVNK QASEQLTLTD VFVFNAADGA
1651 LVEVILGIAY VKIARPSMEK LLARLTEPSW VAGGKTTPQT ATKPAAAPVV
1701 ADHTPRTTES ASTVNGVNLD DRKPEGTALP QEMLSDTEEL RPKAQGQELQ
1751 DMIARVKAVM ADISGLDISE IKDDSNLADL GIDSLVGMEM THEIESTLKV
1801 ELPESEIMSV VDMEGLLQCV AGALGLSMTG ASSDTLTASS DSGINSAKSS
1851 ILSGTSTSTS TGTTDTGSDV GQSMKEPSLM LDTVKKAFAQ TKEATDARIK
1901 AASNQVSYCS TSLPQQNELS VLLTITALEA LGAGFSTARP GSQLTRISHA
1951 PGHEQFVTHL YKEIETATQI IKIDGHGAQA VITRTAVPLP DVESRQVALC
2001 EQMLRGDPEQ VGTMELIKHA GENLHRVLSG ETDGAKVIFG SKTGSKLVSQ
2051 WYAQWPLNRS LIAQMGDFLT AVVAGIQADE DMPFSEINPL RIMETGAGTG
2101 GTTKQIVPLL ARLGLPVVYT FTDLAPSFVA AARKTWGKEY PWMQFRTLDM
2151 EKTPPSVEDG LPLQHFIVSA NAVHATKSIS ATTGNLRKAL RTDGFLLMME
2201 MTRTPFWVDL IFGLFEGWWL FEDGRKHALT HEALWDQELS KVGFGYVDWT
2251 EGMTAESEIQ KIILASADAN TR*WLERVRL PASHTDYHLN QVGVENEARE
2301 LMVADYVSTL TKEFNKTMTQ YTDAGLSLSS RTSQTPMSSQ KRCILITGGT
2351 GGLGAHLVAE AALLPDVNMV ICLNRPNRKQ EARERQLVSL EKKGLILSPE
2401 ALAKITVFET DLSQPGSLGL SDDKYNLLRG NVTHIIHNAW LMHSKWPVRR
2451 FEPQLRIMAH MLNLAADIAT CQRTQGQRQP GPPVSFVFVS SIATVGYHPV
2501 VTNPGNPAVP ETRIPISSVL PTGYGEAKYI CERMLDATLH QYPAQFRASA
2551 VRLGQIAGSE INGHWNSAEH ISFLVKSSQS IGALPALPGP MGWTPADYVA
2601 RGLVEIATQP DNIELYPIYH IENPVRQPWD EALAVLADEM GISSEALPFQ
2651 EWVQTVRDWP RQGDNTAAGA NPAYLLVDFL EDHFLRMSCG GLLLGTAKAR
2701 EHSPSLAGMG PVSDELLRLF VRSWKEVGFL L
```

## 1.4 Cloning of *ASpks1* into pTAex3 and transformation into *A. oryzae*.

205 **1.41 Cloning of PKS.**
An expression plasmid vector was constructed for *ASpks1* based upon pTAex3 and transformed into *A. oryzae* protoplasts. There were no suitable restriction sites before the start codon for *ASpks1* in the KHIII4A2 clone so the 5’ end of the gene was amplified by PCR first and cloned into an entry vector. The forward primer contained the sequence CACC followed by the start codon ATG to allow cloning 210 into pENTR-SD-D-TOPO. The reverse primer had an *Eco*RI site introduced at the 5’ end to create this restriction site into the 3’ end of the PCR product which resulted from its use. An *Xcm*I site was located 684 bp downstream from the start of the gene, and an *Eco*RI site after the stop codon. These two restriction sites allowed the bulk of the gene (8 471 bp) to be obtained by a restriction digest with *Xcm*I and *Eco*RI. This fragment was then inserted into a similarly digested vector. The gene, intact and 215 present in the pENTR-SD-D-TOPO vector, was then recombined with a vector containing the selectable marker, promoter and terminator sequences to allow heterologous expression of the AsPKS1 sequence.

**1.42 Transformation into *A. oryzae*.**
220 A spore suspension (100-200 µl) of the appropriate fungal strain was spread onto an agar plate of the appropriate rich medium. The plate was incubated at 25˚C for 3-5 days. Tween 80 (0.01 %; 10 ml) was added to the agar plate and the spores scraped off with a loop. The liquid was collected and centrifuged (10 000 × g, 5 minutes), the supernatant removed and the crude spore preparation resuspended in water (1 ml). This spore suspension was used to inoculate the appropriate rich growth medium (100 ml) 225 which was then incubated at 25 °C with shaking (240 rpm, 24-48 hours).

Mycelia were collected by filtration through sterile Miracloth, washed twice with 0.8 M sodium chloride (20 ml), centrifuged (10 000 × g, 10 minutes) and the supernatant was discarded. Filter-sterilised protoplasting solution (20 ml, 20 mg/ml Glucanase, 10 mg/ml Driselase; Interspex Products in 0.8 M NaCl) was added to the pellet and resuspended thoroughly by vortexing. The tube was 230 incubated at room temperature, with gentle mixing on a rotator. After one hour, a small sample was analysed microscopically for protoplast formation, if the cells had not formed protoplasts yet, the incubation was continued and the cells checked again at 30 minute intervals until a sufficient proportion of protoplasts had formed.

The protoplasts were released from hyphal strands by gentle pipetting with a large wide-bore 235 pipette (5 ml). The protoplasts were filtered through sterile Miracloth to remove the hyphae. The filtrate was centrifuged (maximum 2 000 × g, 5 minutes) just fast enough to pellet all of the protoplasts, if after centrifugation the solution was cloudy the tube was centrifuged faster and/ or longer. The protoplasts were washed twice with 0.8 M NaCl (20 ml) and once with Solution I (0.8 M NaCl, 10 mM $CaCl_2$, 50 mM Tris-HCl pH 7.5; 20ml), each time the pellet was resuspended by using a 240 wide-bore pipette and centrifuged (2 000 × g, 5 minutes) to remove the supernatant.The concentration of protoplasts was determined by using a haematocytometer (Fisher) and the protoplasts resuspended in Solution I to give a concentration in the range 1-9 x $10^7$ protoplasts/ ml. The protoplasts were stored on ice.

DNA (5-10 µg, 10 µl maximum) to be transformed into the fungus was added to the protoplast 245 suspension (100 µl) and incubated on ice (2 minutes). Solution II (60 % PEG 3350 (Sigma), 10 mM $CaCl_2$, 50 mM Tris-HCl pH 7.5; 1 ml) was added dropwise over 15 minutes, then incubated at room temperature (20 minutes). Plates were poured with Czapek-Dox agar in sorbitol (1 M; 5 ml). Czapek-Dox agar in sorbitol (1 M; 5 ml) was added to the transformation mixture and overlaid onto the prepared plates. The plates were incubated at 25 °C.

250 In the case of the *Aspergillus oryzae* transformations, where selection occurs on arginine-deficient media, the plates were pre-poured with 10 ml agar and the transformed protoplasts plated in

This journal is © The Royal Society of Chemistry 2007

10 ml agar. As the growth was on arginine-deficient media, any transformants were now able to grow. A positive control was made by the addition of 2 % arginine solution (200 µl per 20 ml plate) to untransformed protoplasts and agar, which was mixed and plated; the negative control consisted of untransformed protoplasts on minimal media.

### 1.5 Fermentation and extraction of 3-methylorcinaldehyde.

200µL of the spore suspensions obtained from transformants was inoculated into 100mL starch medium (Starch (20g/L) and polypeptone (10g/L) were added to 900mL of distilled water. To this solution A (50mL; sodium nitrate (40g/L), potassium chloride (40g/L), magnesium sulfate heptahydrate (10g/L), iron sulfate hepthydrate (0.2g/L)) and solution B (50ml; potassium phosphate (20g/l)) were added). The flasks were incubated at 25$^{o}$C with shaking at 150rpm for 6 days.

The culture broth was acidified using 2M hydrochloric acid until the solution reached approx. pH 3.0 and was put back on the shaker for 30mins. The mycelia were ground using a hand blender and then removed by suction filtration through the filter paper and washed with distilled water. The filtrate was washed twice with ethyl acetate. The organic layer was separated from the aqueous phase in a separation funnel and dried over anhydrous magnesium sulfate. The ethyl acetate was evaporated at a reduced pressure by the rotary evaporator. The residue of this crude extract was dissolved in methanol and analysed by LC/MS.

### 1.6 Purification and characterisation of 3-methylorcinaldehyde.

Samples were analysed by HPLC-MS using a Luna 5µm $C_{18}$(2) column (Phenomex, 250 × 4.6mm, 5µm). The program was: 0-5 min (10% B); 5-35 min (gradient to 75% B); 35-37 min (gradient to 90% B); 37-45 min (90% B); 45-50 min (gradient to 10% B). Solvent A: $H_2O$ + 0.1% TFA, solvent B: MeCN + 0.09% TFA. Flow rate 1mL/min, UV 280 nm. Mass data was collected using a Waters Platform II mass spectrometer with an electrospray source operating in positive ionisation mode. 3-Methylorcinaldehyde eluted at 32.36 min, with [M]$H^+$ 167 *m/e*.

3-Methylorcinaldehyde was purified from the crude extract using flash chromarography (30/70 ethylacetate/hexane to 100% ethylacetate). The following physiochemical data identified the *ASpks1* expression product to be 3-methylorcinaldehyde: m.p. 142-152$^{o}$C (lit.,[i] 137-140$^{o}$C). $\delta_H$ (400 MHz, $CDCl_3$) 12.65 (1H, s, 2-O*H*), 10.07 (1H, s, C*H*O), 6.20 (1H, s, 5-*H*), 2.50 (3H, s, 6-C*H*$_3$), 2.08 (3H, s, C*H*$_3$). $\delta_C$ (100 MHz, $CDCl_3$) 193.0 (*C*HO), 164.1 (4-*C*-OH), 161.0 (2-*C*-OH), 141.4 (1-C-*C*H$_3$), 113.3 (6-*C*-CH$_3$), 109.9 (4-*C*H), 108.9 (3-*C*-CH$_3$), 17.9 (6-*C*H$_3$), 6.8 (3-*C*H$_3$). HRMS EI calculated ([$M^+$] for $C_9H_{10}O_3$) 166.0630, found 166.0630

## 2. Crystallographic details for 3-methylorcinaldehyde.

A crystal of 3-methylorsellinaldehyde was prepared by vapour diffusion (MeOH/$CHCl_3$). Crystal data were collected on a Bruker-AXS D8 with a MoK anode generating an X-ray wavelength of 0.71 Å. Data collection and unit cell refinement were performed with SMART v5.628, data integration and reduction were performed using SAINT v7.06A and SHELXTL v6.14 and refinement was performed using SHELXL-97: $C_9H_{10}O_3$, 0.5($H_2O$); M = 175.18; orthorhombic, $a$ = 13.465(3), $b$ = 16.024(3), $c$ = 3.8100(8) Å, $\alpha$ = 90.00, $\beta$ = 90.00, $\gamma$ = 90.00°, $U$ = 822.1(3)$Å^3$, T =100(2) K, space group P-21-21-2, Z = 4, μ(Mo–Kα) = 0.109 $mm^{-1}$, 9320 reflections measured, 1140 unique ($R_{int}$ = 0.0400) which were used in all calculations. The final $R_1$ and $wR_2$ were 0.0478 and 0.1267 (I>2σI).

## 3. Multiple alignment of KS domains from Non-reducing Clade III PKS.

Sequence Analysis of Non-Reducing Clade III PKS peptides in the KS domain. From top: *B. fuckeliana PKS16; C. heterostrophus PKS22; C. heterostrophus PKS23; B. fuckeliana PKS17; C. heterostrophus PKS21; B. fuckeliana PKS18; B. fuckeliana PKS20; Consensus.*

## Notes and References.

i. Y. Jiao, T. Yoshihara, M. Akimoto and A. Ichihara, *Bioscience Biotechnology and Biochemistry,* 1994, **58**, 784-785.