Electronic Supplementary Material (ESI) for Organic & Biomolecular Chemistry.

Supplemental Information

Chorismatases – the family is growing

Mads J. Grüninger,[a,§] Patrick C. F. Buchholz,[b,§] Silja Mordhorst,[a,1] Patrick Strack,[a] Michael Müller,[a] Florian Hubrich,[a,1] Jürgen Pleiss,[b] Jennifer N. Andexer[a,*]

[a] Institute of Pharmaceutical Sciences, University of Freiburg, Albertstr. 25, 79104 Freiburg, Germany

[b] Institute of Biochemistry and Technical Biochemistry, University of Stuttgart, Allmandring 31, 70569 Stuttgart, Germany

[§] equal contributions

[1] present address: Institute of Microbiology, ETH Zuerich, Vladimir-Prelog-Weg 4, 8093 Zuerich, Switzerland

* Correspondence to

Jun.-Prof. Dr. Jennifer N. Andexer

Email: jennifer.andexer@pharmazie.uni-freiburg.de

Tel: +49 761 203 67398

Fax: +49 761 203 6351

Table of Contents

Methods

Sequence data set, phylogenetic analysis, sequence network and correlation analysis: The analysed dataset consisted of 366 (putative) chorismatase protein sequences (Table S3), derived from a BLAST search limited to an expectation value of 10^{-90} with FkbO, Hyg5, and XanB2 as templates. The dataset was clustered to allow for a better overview in a phylogenetic tree (Figure S1) consisting of 78 protein sequences of chorismatase homologues, clustered from Table S3 by the CD-HIT algorithm[1,2] with a word size of 5 and an identity threshold of 0.8. The phylogenetic tree was computed using the web tool from *http://www.phylogeny.fr*, which used the alignment algorithm of MUSCLE (v3.8.31), removed poorly aligned regions by Gblocks (v0.91b) and constructed a maximum likelihood tree by PhyML (v3.1/3.0 aLRT) that was finally visualised using TreeDyn (v198.3).[3–9]

In addition, a protein sequence network for the 78 protein sequences from Figure S1 was constructed by a threshold of 50% sequence similarity from pairwise Needleman-Wunsch alignments (from the implementation of the EMBOSS software suite using gap opening penalty of 10, gap extension penalty of 0.5 and the BLOSUM62 substitution matrix).[10,11] The network was visualised in Cytoscape[12] (v3.7.0) using prefuse force directed layout algorithm with respect to the edge weights of sequence similarity (Figure 5). Family annotations for CH-I, CH-II, CH-III, and CH-IV are coloured as green, blue, dark blue, and yellow, respectively. The network consists of 78 nodes in total, with one isolated node and 77 nodes connected by 1417 edges.

The correlation analysis was carried out using the *GREMLIN* (*Generative REgularized ModeLs of proteINs*) algorithm.[13] For the implementation of the GREMLIN algorithm, MATLAB version R2015a (The Mathworks, Natick, MA, USA) and the MSAvolve toolbox v3.0a were used.[14] Sequences and alignment columns containing more than 25% gaps were discarded,[13] reducing the input alignment to 363 sequences and 309 alignment columns used for correlation analysis. The result of the correlation analysis is presented as Z scores, raw values indicating the strength of the correlation (Table S1, Figure S2). A minimum Z score of 20 was chosen to select correlated pairs for positions 173, 240, 327, and 336, numbered according to Hyg5 from *Streptomyces hygroscopicus*, which was chosen as reference sequence (NCBI accession AAC38060, Uniprot accession O30478, PDB entry 5AG3): If not mentioned otherwise, all position numbers mentioned in this work refer to this sequence.

Cloning, expression, and purification procedures: All genes were synthesised as codon-optimised DNA fragments (GeneArt, Regensburg, Germany). Polymerase chain reaction (PCR) was carried out with Phusion Flash PCR Master Mix, 20 ng of template, and 0.5 µM of the corresponding oligonucleotides (Table S2). The linear DNA fragments were digested with *Nde*I and *Xho*I, and ligated into the vector pET28a(+). Electrically competent DH5α were transformed with this ligation mix without further purification. Plasmids were isolated from grown colonies, digested with *Nde*I and *Xho*I, and analysed via agarose gel electrophoresis. Vectors showing the insert in the corresponding lanes were sequenced by GATC (Konstanz, Germany). For gene expression, electrically competent BL21(DE3) RP pL1sL2 cells[15] were transformed with the vectors. A single colony was used to inoculate an over-night culture; the main cultures were grown in 500 mL medium in 2 L flasks. After reaching an OD_{600} of 0.5, protein production was induced using 0.2 mM isopropyl-β-D-thiogalactopyranosid (IPTG) and the cultures were incubated for 20 h at 24 °C. The cells were harvested using centrifugation (Beckman Coulter, 4 °C, 9000 rpm, 15032 x g). For the protein purification, the cells were disrupted using sonification (Branson, Sonifier, duty cycle 50%, output 80%, 3 times 15 pulses), non-disrupted cells and cell debris were removed by centrifugation and the enzymes were purified using Ni-NTA affinity chromatography. The enzyme was eluted using a step gradient of imidazole, with the fractions 100 and 300 mM containing the major part of the enzyme. These fractions were pooled and concentrated using a 30 kDa concentrator (Pall Corporation, Macrosep Advance Centrifugal Device, 30K). Protein purity was controlled with SDS-PAGE analysis (Figure S3).

Enzyme assays: The production and purification of chorismate was carried out as described previously.[16] Enzyme activity was measured using a continuous, coupled chorismate:NADH:LDH spectrophotometric assay.[17] In brief, chorismate was preincubated in TRIS buffer (100 mM, pH 7) containing lactate dehydrogenase (LDH, 2.5 U/mL) and (NADH, 0.5 mM), until the absorption at 340 nm stabilised. Then, the enzyme (100 µg/mL) was added and the decrease in absorption followed at 340 nm. Chorismate concentrations were between 0.0175 and 2 mM for the determination of kinetic parameters, in the case of *La*CH-II, the chorismate concentration was increased up to 5 mM; kinetic parameters were fitted using the Michaelis Menten model in the Origin Software (OriginPro, 9.1G). For the analysis of product inhibition, 4-HBA (0 – 2 mM) was added to the assay mixtures. Product selectivity was analysed via HPLC/UV as described previously, quantification of product distribution (Figure S6) is based on calibration curves.[18]

DNA and Amino Acid Sequences

DNA sequences (start and stop codon are underlined):

>*A*RapK

ATGGAACGTCTGGTGACCAGCATTACCGCACCGTATTGTCGTTTTGAAAAAGTTGGTAGCAGCGATCTGGAAGCAGATGAAACCCTGATTGG
TGTTGTTGATCATCGTGCAGGTCCTGCCGAAGTTAGTCCGGATGATGGTTATCCGCGTGCAACCGTTCATACCACCACCCGTGATGATGAAA
GCTTTGCGGAAGTTTGGCGTGCACAAGAACGTATTGAAAGCGGTCGTCATGGTGGTATTGCATGGGCACGTACCGAAGATTACCTGTTTGGT
ATTGTTCAGGTTCCGGAAGGTGATCGTTATGCAGCAGCAACCACCACCGCATATACCGCACTGTTTGATCTGACCGCAGGTCTGGGCTATCC
GCATCTGGCACGCACCTGGAATTATGTTAGCGGTATTAATGCAGCAAATGCCGATGGTCTGGAAGTGTATCGTGATTTTTGTGTTGGTCGTG
CCGAAGCACTGGATGCACGTGGTGTTGAACCGGCAAGCATGCCTGCAGCAACCGGTATTGGTGCACATAGCGGTGGTATTACCTGCTGTTTT
CTGGCAGCCCGTGGTGGTGATCGTGTTAATATTGAAAATCCGGCAGTGCTGACCGCACATCGTTATCCGGCACGTTATGGTCCGCGTCCGCC
TGTTTTTGCACGTGCCACCTGGCTGGCACCGCCTGGTGGTGGTGGACGTGGTCGTCTGTTTGTTAGCGCAACCGCAGGTATTCTGGGTCATG
AAACCGTGCATCATGGTGATGTTGCACGTCAGTGTGAAGTTGCACTGGCAAATATTGGTCGTGTTATTGGTGCCGAAAACCTGGCACGTCAT
GGCCTGGATCGTGGTCATACCCTGGCAGAAGTGGATCATCTGAAAGTTTATGTTCGTCATCGTGAAGATGTTGAAACCGTTCGTCGTATTTG
TGCAGATCGTCTGAGCCGTGAAGCAACCTTTGCAGTTCTGCATACCGATATTGCCCGTACCGATCTGCTGGTTGAAATTGAAGGTCTGGTTG
AATAA

>*La*CH-II

ATGCCGAGCAGCGTTTTTACCGGTCTGACCGAAGTTAGTGATGTTGCAGCAACCGGTAATGTTCTGGGTGTTATTAACTATACCACCGAATG
TACCGGTCCGCGTGTTGAAAATGGTCTGCTGACCCTGGATCTGCACATGGCAAGCGATCTGGGTGAAGCATTTGCAGAAGTTTGGACCACCA
GTCGTCCGGCAGAAACCGGTGAACATCATGGTGTTGTTTATGCACATGATGGTGAATATCTGCTGGTTGCAGGTCGTATTGCACCGACAGGT
CGTTATACCGAAGATACCCGTGCAGCCTATGTTGCCGCACTGGATCTGATGGATACCCTGGGCTATAAAAACTGTTTTCGCATGTGGAACTT
CATCAACGATATTAATGCCGATAATACCGAAGGCCTGGAAATCTATCGTGATTTTTGTAAAGGTCGTGCCGAAGCCTTTGAACTGTTTCATT
TTGGTGATAAAGAAGTTCCGAGCGCCACCGGTATTGGTAGCCAGGGTGGTGGTATTGCATTCTATTTTCTGGCAAGCCGTAGCGCAGCACTG
ACCAGCGTTGAAAATAGCAAACAAATGGCAGCATATCATTATCCGCGTCAGTATGGTCCGCGTCCGCCTAAATTTGCACGTGCCACCTATCT
GGCCAGCACCCATACCGATCGTCGTAGCGGTCAGGTTTATGTTGCAGGCACCGCAAGCATTCGTGGTCATGAAACCCTGCATGATGGCGATA
TTGTTGCACAGGTTAAACTGAGCCTGGATAACATTGAACATCTGATTTCCGATGAAAACCTGGCCGATTATGAAATTACCCAGGGCAATAGC
CTGCTGAATCTGGACAACATTAAAGTGTATGTTCGTCACCGTAGCGATATTCCGGTTGTTCGTCGTATGTGTGAAGAACGTTTTAGCCCGAA
TGCACGTGTTCAGTATCTGAATGTTGATGTTTGTCGTGCAGACCTGCTGGTAGAAATTGAAGGTATTGTTGCC

>*Xf*CH-III

ATGAATGCCACCCTGACCACCACCCTGCATATTGATTATGTGCCTCGTGCAACCCTGCCGATGCTGCTGGCAGATCATTGTGTTCTGGCAGT
TTTTGGTTTTGGTGCAGATGCACCGCATGCAACCGATCCGCGTTATGTTCGTGTTCCGCTGGAACCGTATGGTAATACCAGCCTGGAAGTTT
GGCGTAGCAATACCCCGACCCAGGTTGGTTGTAAAGATGGTATTGCATGGGCAAGCAATGATCAGCTGCAGTTTGGCACCATTGCACTGGAA
GAAAATACCCAGCCGATTGATATTTGTAGCGAAGAACTGTATACCCGTCTGACCCATTTTGTTGCAGGTAGCAGCACACCGCATCTGCTGCG
TATTTGGAATTATGTTGATGGCATTACCATTGGCACCGGTGATAATGAACGTTATCGCAAATTTAACATTGGTCGTGCACGTGGTATTAGCG
GTCTGAAAACCGCACAGCTGCCTGCAGCAACCGCAATTGGTTATAGCAATGGTCATCGTATTCTGCATGTTTATTGGCTGGCAAGCGCACAT
CCGGGTACACCGCTGGAAAATCCGCGTCAGATTAGCGCATATTGTTATCCGAAAACCTATGGTCCGCAGCCTCCGAGCTTTGCACGTGCCAT
GCTGCCTCCGCGTAATAGCGCAATGCCTCTGCTGCTGAGCGGCACCGCAGCAATTGTTGGTCATCAGAGCATGCATCCGGGTGATCCGCTGG
CACAGCTGGAAGAGATTTTTGCAAACTTTAATGTGCTGCTGCGTCATGCACATGCACAGCATCCGAAAATTCCGACCCAGTTTGATCTGAAA
ACACGCCTGAAAGTTTATGTGCGTGAGCAGATGGATCTGCCGCATATTGCAACCGCACTGCAAGAACGTCTGGGTGCAATTCCTCATCTGCT
GCTGCATGGCACCATTTGTCGTAACGATCTGCTGGTTGAAATTGATGGTTATACAGGC

>*Sm*CH-IV

ATGAATGCCACCCTGACCACCACCCTGCATATTGATTATGTGCCTCGTGCAACCCTGCCGATGCTGCTGGCAGATCATTGTGTTCTGGCAGT
TTTTGGTTTTGGTGCAGATGCACCGCATGCAACCGATCCGCGTTATGTTCGTGTTCCGCTGGAACCGTATGGTAATACCAGCCTGGAAGTTT
GGCGTAGCAATACCCCGACCCAGGTTGGTTGTAAAGATGGTATTGCATGGGCAAGCAATGATCAGCTGCAGTTTGGCACCATTGCACTGGAA
GAAAATACCCAGCCGATTGATATTTGTAGCGAAGAACTGTATACCCGTCTGACCCATTTTGTTGCAGGTAGCAGCACACCGCATCTGCTGCG
TATTTGGAATTATGTTGATGGCATTACCATTGGCACCGGTGATAATGAACGTTATCGCAAATTTAACATTGGTCGTGCACGTGGTATTAGCG
GTCTGAAAACCGCACAGCTGCCTGCAGCAACCTGCATTGGTTATAGCAATGGTCATCGTATTCTGCATGTTTATTGGCTGGCAAGCGCACAT
CCGGGTACACCGCTGGAAAATCCGCGTCAGATTAGCGCATATTGTTATCCGAAAACCTATGGTCCGCAGCCTCCGAGCTTTGCACGTGCCAT
GCTGCCTCCGCGTAATAGCGCAATGCCTCTGCTGCTGAGCGGCACCGCAGCAATTGTTGGTCATCAGAGCATGCATCCGGGTGATCCGCTGG
CACAGCTGGAAGAGATTTTTGCAAACTTTAATGTGCTGCTGCGTCATGCACATGCACAGCATCCGAAAATTCCGACCCAGTTTGATCTGAAA
ACACGCCTGAAAGTTTATGTGCGTGAGCAGATGGATCTGCCGCATATTGCAACCGCACTGCAAGAACGTCTGGGTGCAATTCCTCATCTGCT
GCTGCATGGCACCATTTGTCGTAACGATCTGCTGGTTGAAATTGAAGGTTATACAGGC

>*Sp*CH-IV

ATGAGCCATCCGATTCATCTGCAGTGTAGCTTTGTTGATGCCGGTGCAGTTGCAGAACAGCTGCGTGATCCGCGTGTTCTGGCAGCAGTTGA
TTTTAGCCTGGAACCGCTGGATGCAGAAGCAGCAGCAGATCCTCGTCGTCTGCGTGTTCCGCTGGAAGCAGTTGGTCCGGGTCGTATTGAAC
TGTGGCGTGGTGATACACCGGCAACACATGGTCGTGAAGGTGATATTGCATGGGCAGAAAATGGTGAGCTGCAGTTTGGTGCACTGGTTCTG
CATGAAAGCGCAGATCTGGATCTGGAAACCGCAAGCGCACAGGCCTATGCACAGATGAACCGTTTTATTGCAGCACGTGGTTATCCGCATGT
GCTGCGTATTTGGAATTATCTGGATGGTCTGACCGAAGGTGAAGATGATGCAGAACGTTATCGTCGTTTTTGTGTTGGTCGTGTTCGTGGTC
TGGAACGTGTTGAAGAAGCACGTCTGCCTGCAGCAACCTGTATTGGTAGCTTTGGTGGTCCGCGTCAGCTGCTGGTTTATTGGCTGGCAGCA
CGTAGTCCGGGTGTTGCACTGGAAAATCCTCGTCAGGTTAGCGCATATCGTTATCCACGTCAGTATGGTCCGCAGAGCCCGAGCTTTGCACG
TGCCATGCTGCCTCCGCCTGGTAGCCAGGCACCGCTGATGCTGAGCGGCACCGCAAGCATTGTTGGTCATGCAACCGCACATCAGGGTGATG
TTCGTGCACAGCTGGATGAAATTCTGGTTAATATTGAAGCACTGCGTGGTGCAGCAGCCGAACGTAGCGGTGAAATGCCTGCAGGTATTGAT
GATGCGGGTACACTGCTGAAAGTTTATGTTCGTGATCGTGAAGCCCTGCCGCAGGTTGCAGCAGCACTGGATGCACGTTTTGGCACACGCGT
TCCTCGTCTGCTGTTACATGCCGAAGTTTGTCGTCGTGAACTGGCAGTTGAAATTGAAGGTGTTCTGGGTAATCCGGCAAGCTAA

Protein sequences:

>*A*RapK

MGSSHHHHHHSSGLVPRGSHMMERLVTSITAPYCRFEKVGSSDLEADETLIGVVDHRAGPAEVSPDDGYPRATVHTTTRDDESFAEVWRAQE
RIESGRHGGIAWARTEDYLFGIVQVPEGDRYAAATTTAYTALFDLTAGLGYPHLARTWNYVSGINAANADGLEVYRDFCVGRAEALDARGVE
PASMPAATGIGAHSGGITCCFLAARGGDRVNIENPAVLTAHRYPARYGPRPPVFARATWLAPPGGGGRGRLFVSATAGILGHETVHHGDVAR
QCEVALANIGRVIGAENLARHGLDRGHTLAEVDHLKVYVRHREDVETVRRICADRLSREATFAVLHTDIARTDLLVEIEGLVE

>*La*CH-II

MGSSHHHHHHSSGLVPRGSHMMPSSVFTGLTEVSDVAATGNVLGVINYTTECTGPRVENGLLTLDLHMASDLGEAFAEVWTTSRPAETGEHH
GVVYAHDGEYLLVAGRIAPTGRYTEDTRAAYVAALDLMDTLGYKNCFRMWNFINDINADNTEGLEIYRDFCKGRAEAFELFHFGDKEVPSAT
GIGSQGGGIAFYFLASRSAALTSVENSKQMAAYHYPRQYGPRPPKFARATYLASTHTDRRSGQVYVAGTASIRGHETLHDGDIVAQVKLSLD
NIEHLISDENLADYEITQGNSLLNLDNIKVYVRHRSDIPVVRRMCEERFSPNARVQYLNVDVCRADLLVEIEGIVALEHHHHHH

>*Xf*CH-III

MGSSHHHHHHSSGLVPRGSHMMNATLTTTLHIDYVPRATLPMLLADHCVLAVFGFGADAPHATDPRYVRVPLEPYGNTSLEVWRSNTPTQVG
CKDGIAWASNDQLQFGTIALEENTQPIDICSEELYTRLTHFVAGSSTPHLLRIWNYVDGITIGTGDNERYRKFNIGRARGISGLKTAQLPAA
TAIGYSNGHRILHVYWLASAHPGTPLENPRQISAYCYPKTYGPQPPSFARAMLPPRNSAMPLLLSGTAAIVGHQSMHPGDPLAQLEEIFANF
NVLLRHAHAQHPKIPTQFDLKTRLKVYVREQMDLPHIATALQERLGAIPHLLLHGTICRNDLLVEIDGYTGLEHHHHHH

>*Sm*CH-IV

MGSSHHHHHHSSGLVPRGSHMMFRVAYVPATQREAALADPRTLAVFGFGLPRARLDDPRWLQVPLAEQGPAQIEVWRGGTPVEHGVTDGVRW
SHNDSLLFGVLEIEETDSDIESAAAEAYARMSAFLSRCGYPHLLRTWNYLDAVTEGDGDQERYRRFCVGRVRGLRELDEAALPAATCIGRFD
GERRLQVYWLAAREPGQPLENPRQVSAFRYPRQYGPQSPSFSRALLPPLATGLPLLQSGTAAIVGHVSQHTGAVDEQLRETLTNLQSLVDTA
RGQRPSLSPVLGSGSVLKVYVRRAEDMGLVAAQMAALPTAPPFVVLHAEVCRAELLVEIEGLHGLEHHHHHH

>*Sp*CH-IV

MGSSHHHHHHSSGLVPRGSHMMSHPIHLQCSFVDAGAVAEQLRDPRVLAAVDFSLEPLDAEAAADPRRLRVPLEAVGPGRIELWRGDTPATH
GREGDIAWAENGELQFGALVLHESADLDLETASAQAYAQMNRFIAARGYPHVLRIWNYLDGLTEGEDDAERYRRFCVGRVRGLERVEARLPA
ATCIGSFGGPRQLLVYWLAARSPGVALENPRQVSAYRYPRQYGPQSPSFARAMLPPPGSQAPLMLSGTASIVGHATAHQGDVRAQLDEILVN
IEALRGAAAERSGEMPAGIDDAGTLLKVYVRDREALPQVAAALDARFGTRVPRLLLHAEVCRRELAVEIEGVLGNPAS

Results and Discussion

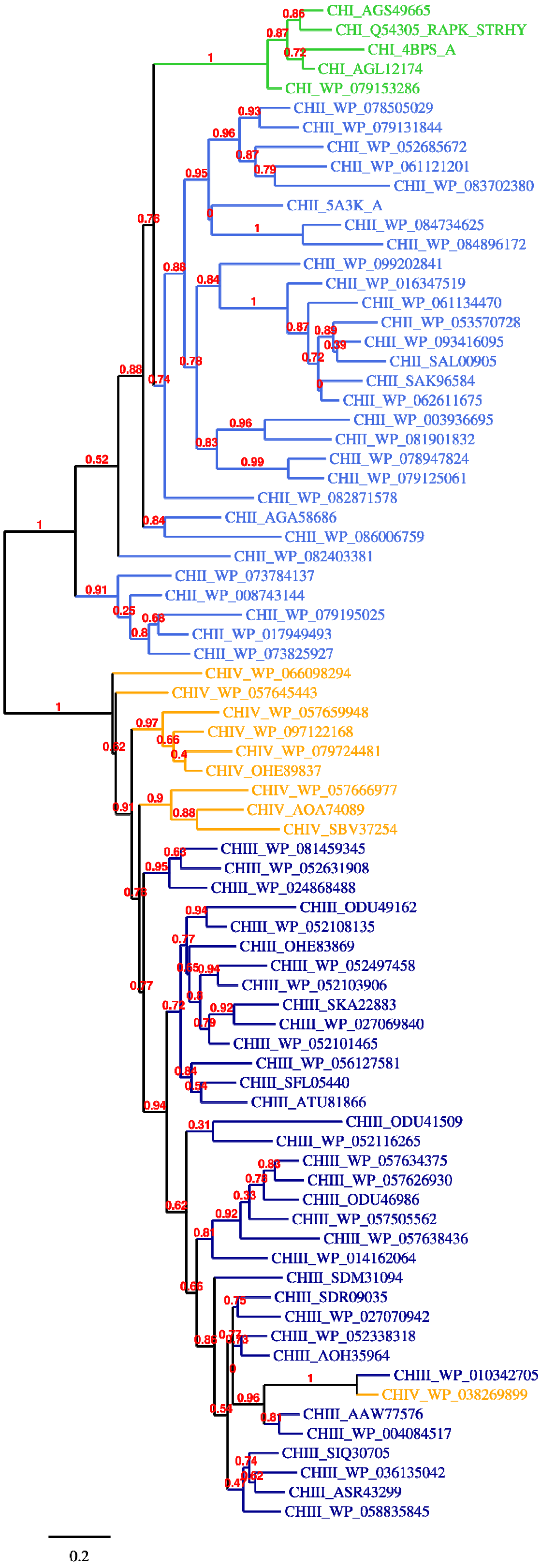


Figure S1. Phylogenetic tree of 78 protein sequences of chorismatase homologues. Red numbers indicate bootstrap values as confidence indices.[8]

Table S1. Position pairs with Z scores ≥ 20 for positions 173, 240, 327, and 336 from the correlation analysis of 366 chorismatase homologues (Table S3).

Reference position	Correlated position	Z score
173	136	22
173	159	21
173	**336**	**23**
240	201	29
240	**327**	**29**
327	201	29
327	**240**	**29**
336	**173**	**23**
336	266	22

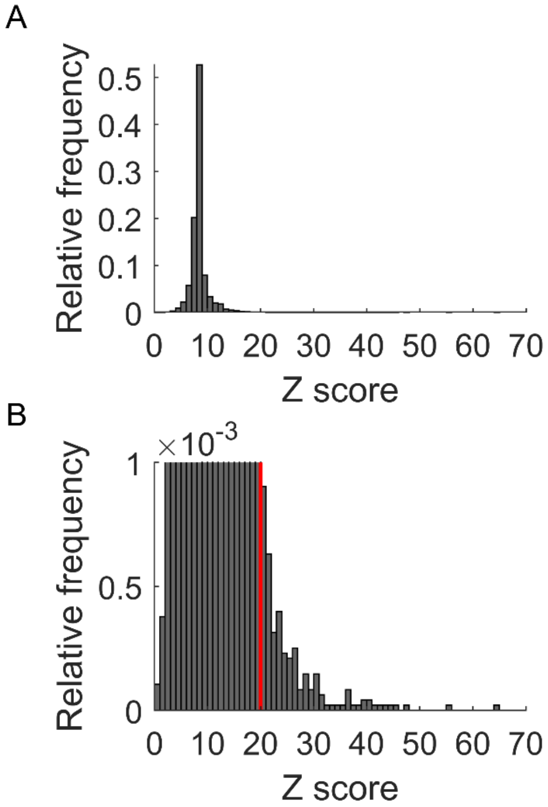


Figure S2. Histogram of Z scores for the correlation analysis of 366 chorismatase homologues (Table S3) (A) with a detail view highlighting the minimum Z score of 20 that was set to choose correlated positions (B).

Table S2. Oligonucleotides used in this study. Restriction sites are formatted bold.

Oligonucleotide	Sequence
*A*RapK_fwd	5′-TATATATA**CATATG**ATGGAACGTCTGGTGACC-3′
*A*RapK_rev	5′-TATATATA**CTCGAG**TTATTCAACCAGACCTTC-3′
*La*CH-II_fwd	5′-GCGGCAGC**CATATG**ATGCCGAGCAG-3′
*La*CH-II_rev	5′-GGTGGTGGTG**CTCGAG**GGCAACAATACC-3′
*Xf*CH-III_fwd	5′-CGCGGCAGC**CATATG**ATGAATGCCACCCTGACC-3′
*Xf*CH-III_rev	5′-GTGGTGGTG**CTCGAG**GCCTGTATAACCATCAATTTCAAC-3′
*Sm*CH-IV_fwd	5′-TATATATA**CATATG**ATGTTCCG-3′
*Sm*CH-IV_rev	5′-TATATATA**CTCGAG**GCCATGCAG-3′
*Sp*CH-IV_fwd	5′-CCTGGTGCCGCGCGGCAGC**CATATG**ATGAGCCATCCGATTCATTC-3′
*Sp*CH-IV_rev	5′-CAGTGGTGGTGGTGGTGGTG**CTCGAG**TTAGCTTGCCGGATTACC-3′

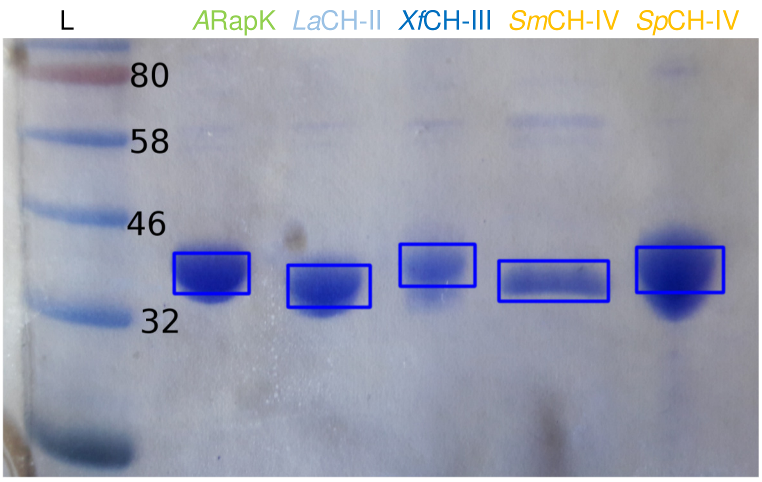
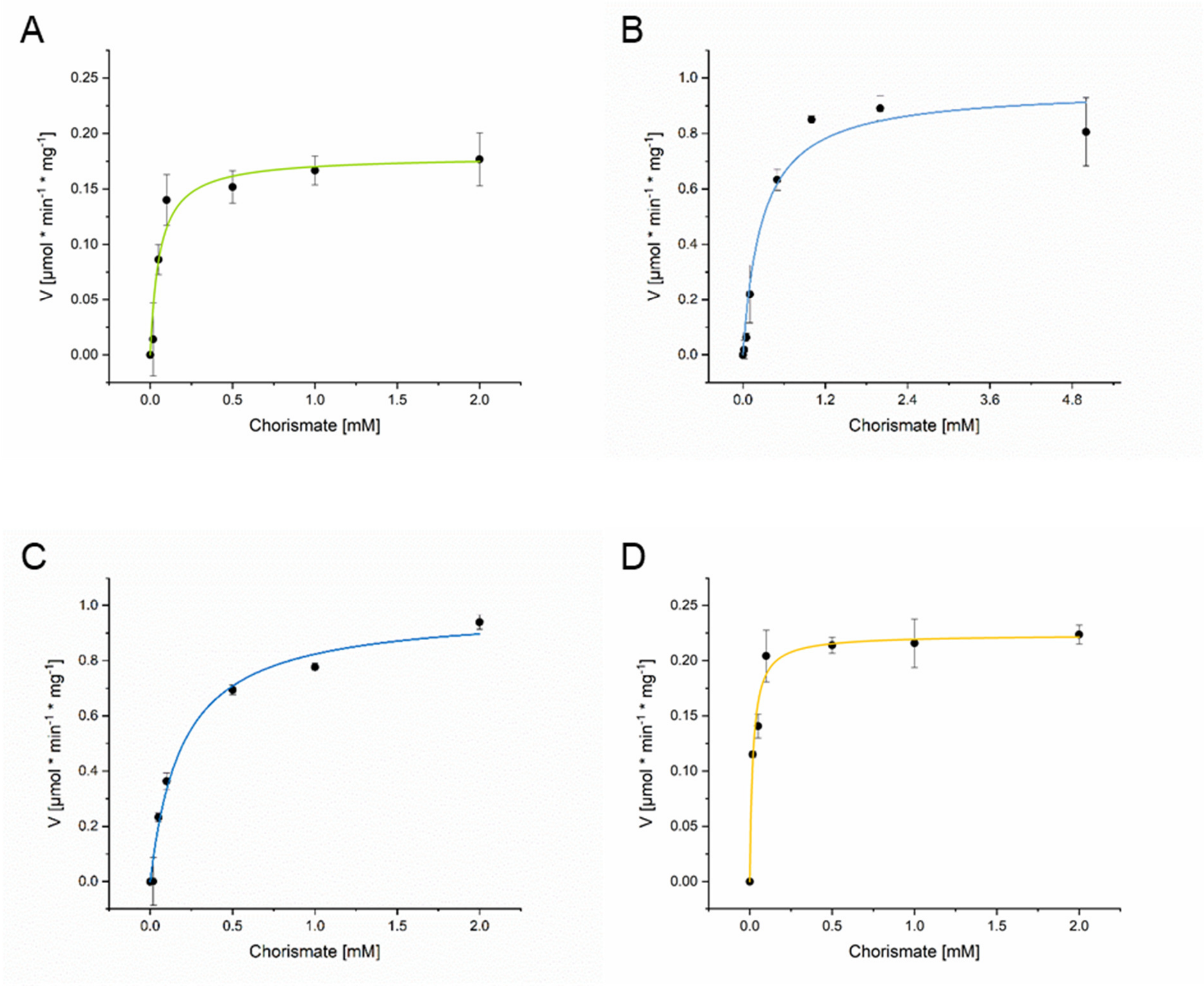
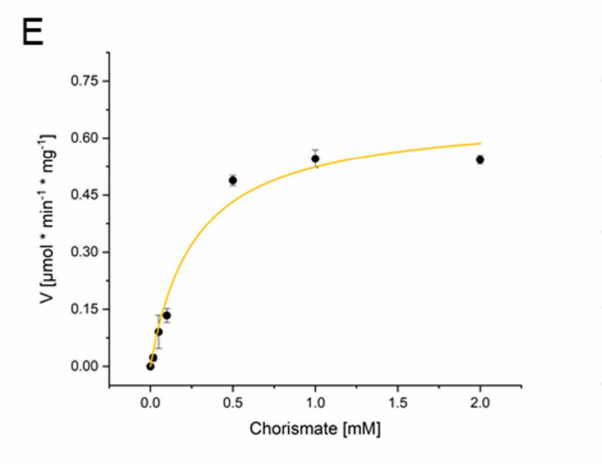


Figure S3. SDS-PAGE analysis of purified, new chorismatases.
L – ladder (Prestained ColorPlus Protein Marker, New England Biolabs, Ipswich, USA), *A*RapK (39.5 kDa), *La*CH-II (41.7 kDa), *Xf*CH-III (41.2 kDa), *Sm*CH-IV (39.8 kDa), and *Sp*CH-IV (40.2 kDa).





Enzyme	**K_M [µM]**	**v_{max} [µmol * min^{-1} * mg^{-1}]**
*A*RapK	56.1 ± 17.9	0.180 ± 0.013
*La*CH-II	279 ± 83.7	0.963 ± 0.068
*Xf*CH-III	195 ± 41.5	0.984 ± 0.054
*Sm*CH-IV	19.1 ± 3.9	0.223 ± 0.008
*Sp*CH-IV	270 ± 76.8	0.665 ± 0.053

Figure S4. Michaelis-Menten kinetics of the new chorismatases' subfamily members. These were determined using the NADH:LDH:chorismate assay. (A) *A*RapK, (B) *La*CH-II, (C) *Xf*CH-III, (D) *Sm*CH-IV, and (E) *Sp*CH-IV.

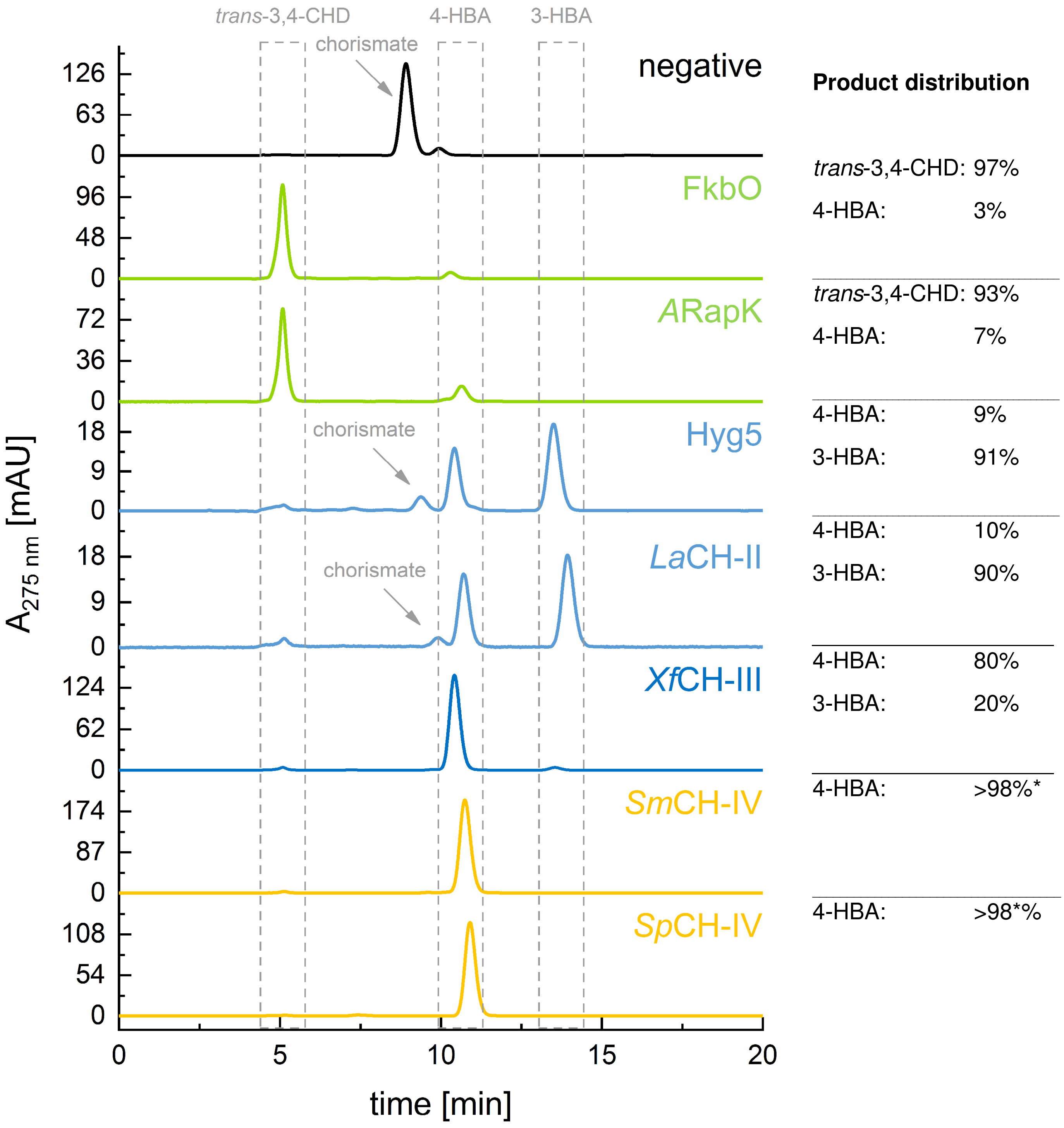


Figure S5. HPLC chromatograms of the wild type chorismatases. Retention times: 5.0 min *trans*-3,4-CHD, 9.0 min chorismate, 10.5 min 4-HBA, and 14.0 min 3-HBA. Assay conditions: 1 mM chorismate, 10 µg of the corresponding chorismatase, 100 mM Tris-HCl pH 7, 20 h, room temperature. The extinction coefficient of 4-HBA is significantly higher than the one of 3-HBA, therefore the 4-HBA peaks appear to be much bigger. As we experienced earlier, the 3,4-*trans*-CHD peak appears just after the dead volume of the column, and the small odd-shaped peaks are most likely not 3,4-*trans*-CHD (shown by LC-MS in earlier tests). Quantifications are based on calibration curves (*trans*-3,4-CHD: y=2318.9x; 4-HBA: y=6866.8x; 3-HBA: y=1014.4x; measurements were carried out in triplicates, R^2 was at least 0.99 in all cases). * detection limit for 3-HBA <2%.

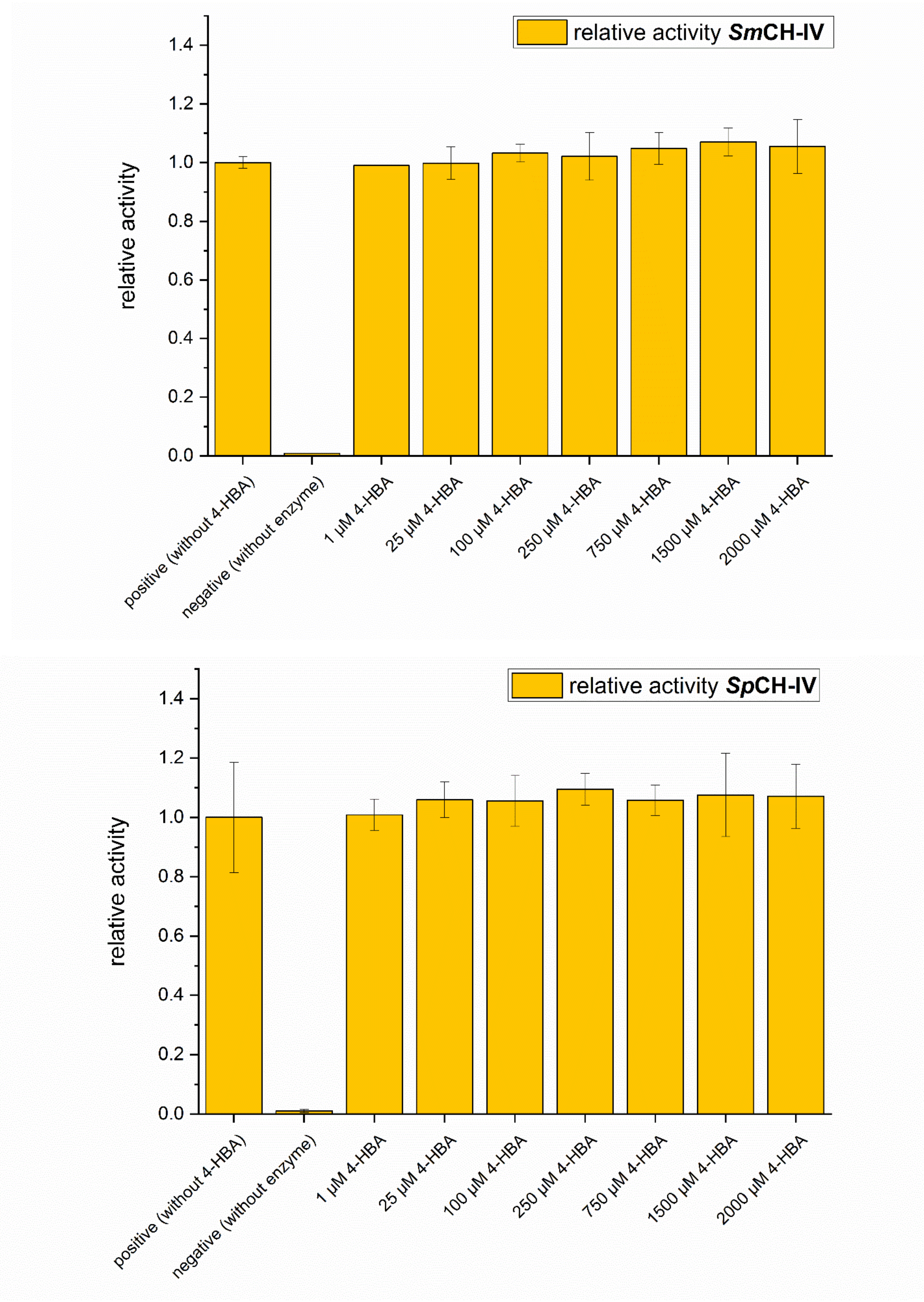


Figure S6. Assay for product inhibition of CH-IV chorismatases by 4-HBA.

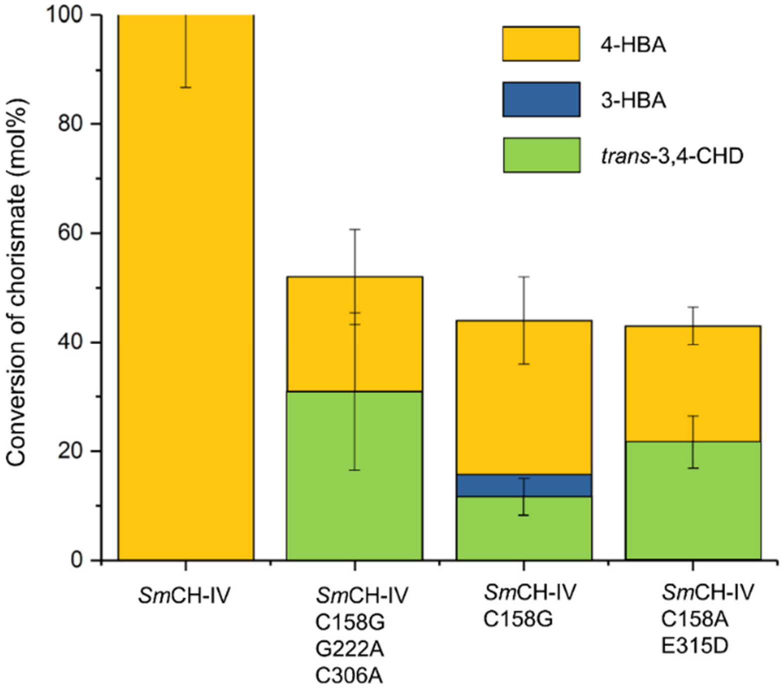


Figure S7. Product range of *Sm*CH-IV variants with altered active site residues (*Sm*CH-IV C158G G222A C306A: CH-I-like; *Sm*CH-IV C158G: CH-II-like; *Sm*CH-IV C158A E315D: CH-III-like). The variants are characterised by a loss of activity (reactions were run for 20 h) and a loss of product selectivity.

Table S3. NCBI accessions and source names for the 366 chorismatase homologues used for correlation analysis, with family assignment based on the amino acid positions from Table 1. The reference sequence used for correlation analysis is highlighted.

Family	Accession (NCBI)	Organism
CH-I	ATL88523	*Streptomyces malaysiensis*
CH-I	SCG08028	*Streptomyces* sp. MnatMP-M27
CH-I	WP_079153286	*Streptomyces* sp. SPMA113
CH-I	WP_093703679	*Streptomyces* sp. MnatMP-M27
CH-I	WP_099016612	*Streptomyces malaysiensis*
CH-I	WP_015621035	*Actinoplanes* sp. N902-109
CH-I	AGL12174	*Actinoplanes* sp. N902-109
CH-I	Q54305	*Streptomyces hygroscopicus* (strain ATCC 29253)
CH-I	AGS49665	uncultured bacterium esnapd13
CH-I	AGP59510	*Streptomyces rapamycinicus* NRRL 5491
CH-I	WP_044578214	*Streptomyces iranensis*
CH-I	OON71283	*Streptomyces tsukubensis*
CH-I	WP_006350824	*Streptomyces tsukubensis*
CH-I	WP_077974555	*Streptomyces tsukubensis*
CH-I	AAC68817	*Streptomyces* sp. MA6548
CH-I	WP_055547736	*Streptomyces kanamyceticus*
CH-I	ADX99525	*Streptomyces* sp. MJM7001
CH-I	4BPS_A	*Streptomyces hygroscopicus*
CH-I	Q9KID9	*Streptomyces hygroscopicus*
CH-II	SAK96584	*Caballeronia hypogeia*
CH-II	WP_061125235	*Caballeronia catudaia*
CH-II	WP_061172366	*Caballeronia hypogeia*
CH-II	WP_061148471	*Caballeronia arvi*
CH-II	SAL05528	*Caballeronia calidae*
CH-II	WP_062611675	*Caballeronia calidae*
CH-II	OKJ48650	*Streptomyces* sp. CB02009
CH-II	WP_008347030	*Burkholderiaceae*
CH-II	WP_061163222	*Caballeronia temeraria*
CH-II	WP_073918801	*Streptomyces* sp. CB02009
CH-II	WP_081916102	*Saccharothrix* sp. NRRL B-16314
CH-II	WP_073892363	*Saccharothrix* sp. CB00851
CH-II	SAL00905	*Caballeronia ptereochthonis*
CH-II	WP_016347519	*Burkholderia* sp. RPE64
CH-II	WP_061121201	*Caballeronia turbans*
CH-II	WP_087049627	*Caballeronia ptereochthonis*
CH-II	OEV13029	*Streptomyces nanshensis*
CH-II	BAG16282	*Nocardia terpenica*
CH-II	WP_062262925	*Caballeronia megalochromosomata*
CH-II	OKJ97282	*Streptomyces* sp. CB02400
CH-II	WP_079195025	*Streptomyces* sp. CB02400
CH-II	ANZ43622	*Lentzea guizhouensis*
CH-II	WP_035510255	*Caballeronia jiangsuensis*
CH-II	WP_061134470	*Caballeronia fortuita*
CH-II	KOX28756	*Saccharothrix* sp. NRRL B-16348

CH-II	WP_082403381	*Saccharothrix* sp. NRRL B-16348
CH-II	WP_083883356	*Thermobacillus composti*
CH-II	WP_083907593	*Pseudonocardia* sp. P2
CH-II	KJK33239	*Lechevalieria aerocolonigenes*
CH-II	WP_086006759	*Nocardia abscessus*
CH-II	WP_081237158	*Streptomyces viridosporus*
CH-II	OLL89571	*Pseudonocardia* sp. Ae406_Ps2
CH-II	WP_083695842	*Pseudonocardia*
CH-II	WP_083702380	*Pseudonocardia* sp. Ae717_Ps2
CH-II	WP_061177983	*Caballeronia pedi*
CH-II	WP_077008874	*Saccharothrix* sp. ALI-22-I
CH-II	WP_085572016	*Streptomyces* sp. 13-12-16
CH-II	AGA58686	*Thermobacillus composti* KWC4
CH-II	WP_053570728	*Caballeronia cordobensis*
CH-II	OLF14593	*Actinophytocola xinjiangensis*
CH-II	WP_008743144	*Streptomyces* sp. Mg1
CH-II	WP_017949493	*Streptomyces* sp. CNS615
CH-II	WP_027734760	*Streptomyces* sp. CNR698
CH-II	WP_073797090	*Streptomyces* sp. CB03578
CH-II	WP_073825927	*Streptomyces* sp. CB02130
CH-II	WP_003936695	*Rhodococcus ruber*
CH-II	KZM74098	*Nocardia terpenica*
CH-II	OKH96755	*Streptomyces uncialis*
CH-II	SCK62708	*Streptomyces* sp. AmelKG-E11A
CH-II	WP_073784137	*Streptomyces uncialis*
CH-II	WP_099279418	*Streptomyces* sp. AmelKG-E11A
CH-II	WP_099202841	*Streptomyces cinnamoneus*
CH-II	WP_084896172	*Streptomyces* sp. CB03238
CH-II	WP_079125061	*Streptomyces lushanensis*
CH-II	WP_082871578	*Nocardia terpenica*
CH-II	WP_098698619	*Nocardia terpenica*
CH-II	WP_093416095	*Saccharopolyspora flava*
CH-II	WP_095583454	*Streptomyces albireticuli*
CH-II	WP_079131844	*Streptomyces nanshensis*
CH-II	WP_087926735	*Streptomyces albireticuli*
CH-II	WP_089951167	*Lechevalieria xinjiangensis*
CH-II	WP_083268678	*Lentzea guizhouensis*
CH-II	WP_078947824	*Streptomyces griseus*
CH-II	WP_081901832	*Lechevalieria aerocolonigenes*
CH-II	WP_084734625	*Actinophytocola xinjiangensis*
CH-II	WP_078505029	*Streptomyces* sp. TAA204
CH-II	WP_052685672	*Lechevalieria aerocolonigenes*
CH-II	WP_093588251	*Lentzea waywayandensis*
CH-II	AGZ78453	*Streptomyces* sp. RJA2928
CH-II	AQA10796	*Streptomyces autolyticus*
CH-II	AGP61380	*Streptomyces rapamycinicus* NRRL 5491
CH-II	WP_079256907	*Streptomyces autolyticus*
CH-II	WP_099013050	*Streptomyces malaysiensis*

CH-II	WP_093705489	*Streptomyces* sp. MnatMP-M27
CH-II	WP_079153844	*Streptomyces* sp. SPMA113
CH-II	**O30478**	***Streptomyces hygroscopicus***
CH-II	5A3K	*Streptomyces hygroscopicus*
CH-III	AAF83579	*Xylella fastidiosa* 9a5c
CH-III	KNE32666	*Xanthomonas campestris* pv. *musacearum* NCPPB 2251
CH-III	KTF29665	*Xanthomonas vesicatoria*
CH-III	KWS05912	*Lysobacter capsici* AZ78
CH-III	ODU41509	*Xanthomonadaceae* bacterium SCN 69-123
CH-III	OLH49107	*Xanthomonas oryzae* pv. *oryzae*
CH-III	WP_046936481	*Xanthomonas gardneri*
CH-III	WP_058569499	*Xylella fastidiosa*
CH-III	WP_065033426	*Xanthomonas arboricola*
CH-III	WP_065624965	*Xanthomonas euvesicatoria*
CH-III	WP_081422697	*Xanthomonas cannabis*
CH-III	AAM38938	*Xanthomonas axonopodis* pv. *citri* str. 306
CH-III	AAM43235	*Xanthomonas campestris* pv. *campestris* str. ATCC 33913
CH-III	AAO29720	*Xylella fastidiosa* Temecula1
CH-III	AAW77576	*Xanthomonas oryzae* pv. *oryzae* KACC 10331
CH-III	ABA27135	*Xanthomonas campestris* pv. *campestris*
CH-III	ADV28623	*Pseudoxanthomonas suwonensis* 11-1
CH-III	AEL05177	*Xanthomonas campestris* pv. *raphani* 756C
CH-III	AEO44208	*Xanthomonas axonopodis* pv. *citrumelo* F1
CH-III	AKC88311	*Pseudoxanthomonas suwonensis*
CH-III	AKS17932	*Xanthomonas campestris* pv. *campestris*
CH-III	ALN64929	*Lysobacter antibioticus*
CH-III	ALN93997	*Lysobacter gummosus*
CH-III	AOH35964	*Luteimonas* sp. JM171
CH-III	ASR43299	*Xanthomonas citri* pv. *mangiferaeindicae*
CH-III	ATS36839	*Xanthomonas citri* pv. *phaseoli* var. *fuscans*
CH-III	ATU81866	*Lysobacter enzymogenes*
CH-III	CAJ25922	*Xanthomonas campestris* pv. *vesicatoria* str. 85-10
CH-III	CAP53576	*Xanthomonas campestris* pv. *campestris*
CH-III	CCF67357	*Xanthomonas citri* pv. *punicae* str. LMG 859
CH-III	CCG37781	*Xanthomonas citri* pv. *mangiferaeindicae* LMG 941
CH-III	CDF63373	*Xanthomonas fuscans* subsp. *fuscans*
CH-III	CDN20012	*Xanthomonas campestris* pv. *viticola*
CH-III	CEM60102	*Xanthomonas campestris* pv. *campestris*
CH-III	EFF43043	*Xanthomonas fuscans* subsp. *aurantifolii* str. ICPB 11122
CH-III	EGD08322	*Xanthomonas vesicatoria* ATCC 35937
CH-III	KFA00015	*Xanthomonas vasicola* pv. *vasculorum* NCPPB 890
CH-III	KFA23887	*Xanthomonas vasicola* pv. *vasculorum* NCPPB 1326
CH-III	KGE50779	*Xanthomonas axonopodis* pv. *vasculorum*
CH-III	KGK55980	*Xanthomonas cannabis* pv. *phaseoli*
CH-III	KGM51312	*Lysobacter concretionis* Ko07 = DSM 16239
CH-III	KGM54234	*Lysobacter daejeonensis* GH1-9

CH-III	KGM54925	*Lysobacter arseniciresistens* ZS79
CH-III	KGO98154	*Lysobacter defluvii* IMMIB APB-9 = DSM 18482
CH-III	KGT52636	*Xanthomonas phaseoli* pv*. phaseoli*
CH-III	KGU57311	*Xanthomonas phaseoli* pv*. phaseoli*
CH-III	KHL52813	*Xanthomonas cannabis* pv*. cannabis*
CH-III	KHM91105	*Xanthomonas vesicatoria*
CH-III	KIQ25034	*Xanthomonas campestris*
CH-III	KLD75768	*Xanthomonas hyacinthi* DSM 19077
CH-III	KOR45575	*Xanthomonas oryzae*
CH-III	KRG60673	*Stenotrophomonas nitritireducens*
CH-III	KRG70824	*Stenotrophomonas terrae*
CH-III	KRG75633	*Stenotrophomonas ginsengisoli*
CH-III	KRG87063	*Stenotrophomonas acidaminiphila*
CH-III	KRG87571	*Stenotrophomonas daejeonensis*
CH-III	OAG68175	*Xanthomonas floridensis*
CH-III	OAX89210	*Xanthomonas nasturtii*
CH-III	ODU46986	*Xanthomonadaceae bacterium* SCN 69-123
CH-III	ODU49162	*Xanthomonadaceae bacterium* SCN 69-48
CH-III	ODU49180	*Xanthomonadaceae bacterium* SCN 69-48
CH-III	OHE83869	*Xanthomonadales bacterium* RIFOXYA1_FULL_69_10
CH-III	OHE88285	*Xanthomonadales bacterium* RIFOXYD1_FULL_69_11
CH-III	OHX25208	*Xanthomonas alfalfae*
CH-III	OOW65189	*Xanthomonas axonopodis* pv*. melhusii*
CH-III	OOW71330	*Pseudomonas cissicola*
CH-III	OOW74886	*Xanthomonas axonopodis* pv*. martyniicola*
CH-III	OOW82585	*Xanthomonas axonopodis* pv*. clitoriae*
CH-III	OOW95625	*Xanthomonas campestris* pv*. vitistrifoliae*
CH-III	OQP79591	*Xanthomonas phaseoli* pv*. syngonii* LMG 9055
CH-III	OQP81872	*Xanthomonas phaseoli* pv*. dieffenbachiae*
CH-III	OWB23348	*Xanthomonas oryzae* pv*. oryzicola*
CH-III	SBV50289	*Xanthomonas bromi*
CH-III	SDM31094	*Oryzisolibacter propanilivorax*
CH-III	SDR09035	*Pseudoxanthomonas* sp*.* CF125
CH-III	SEM39845	*Pseudoxanthomonas* sp*.* GM95
CH-III	SFL05440	*Lysobacter* sp. cf310
CH-III	SIQ30705	*Luteimonas tolerans*
CH-III	SKA22883	*Lysobacter spongiicola* DSM 21749
CH-III	SMQ93619	*Xanthomonas fragariae*
CH-III	WP_002809632	*Xanthomonas fragariae*
CH-III	WP_003470211	*Xanthomonas translucens*
CH-III	WP_003476651	*Xanthomonas translucens*
CH-III	WP_004084517	*Xylella fastidiosa*
CH-III	WP_004090449	*Xylella fastidiosa*
CH-III	WP_006451939	*Xanthomonas gardneri*
CH-III	WP_009572838	*Xanthomonas translucens*
CH-III	WP_010342705	*Xanthomonas sacchari*
CH-III	WP_010369742	*Xanthomonas*
CH-III	WP_011039085	*Xanthomonas campestris*

CH-III	WP_011052731	*Xanthomonas*
CH-III	WP_011260628	*Xanthomonas oryzae*
CH-III	WP_011409665	*Xanthomonas oryzae*
CH-III	WP_012914689	*Xanthomonas albilineans*
CH-III	WP_014162064	*Pseudoxanthomonas spadix*
CH-III	WP_014501624	*Xanthomonas oryzae*
CH-III	WP_016902830	*Xanthomonas arboricola*
CH-III	WP_017112524	*Xanthomonas vasicola*
CH-III	WP_017908313	*Xanthomonas* sp. SHU 199
CH-III	WP_017911664	*Xanthomonas* sp. SHU 166
CH-III	WP_017916745	*Xanthomonas* sp. SHU 308
CH-III	WP_019300736	*Xanthomonas oryzae*
CH-III	WP_019397292	*Pseudoxanthomonas*
CH-III	WP_020851227	*Xylella fastidiosa*
CH-III	WP_022971487	*Xanthomonas maliensis*
CH-III	WP_023905423	*Xanthomonas hortorum*
CH-III	WP_023906989	*Xylella fastidiosa*
CH-III	WP_023908091	*Xylella fastidiosa*
CH-III	WP_024868488	*Pseudoxanthomonas suwonensis*
CH-III	WP_024937538	*Xanthomonas arboricola*
CH-III	WP_027069840	*Lysobacter defluvii*
CH-III	WP_027070942	*Luteimonas* sp. J29
CH-III	WP_031371267	*Lysobacter antibioticus*
CH-III	WP_036108069	*Lysobacter capsici*
CH-III	WP_036135042	*Luteimonas mephitis*
CH-III	WP_038211261	*Xylella fastidiosa*
CH-III	WP_039008463	*Xanthomonas translucens*
CH-III	WP_039512735	*Xanthomonas arboricola*
CH-III	WP_039521164	*Xanthomonas arboricola*
CH-III	WP_039524621	*Xanthomonas arboricola*
CH-III	WP_039814860	*Xanthomonas arboricola*
CH-III	WP_042462926	*Xylella fastidiosa*
CH-III	WP_043093717	*Xanthomonas sacchari*
CH-III	WP_045726593	*Xanthomonas* sp. GPE 39
CH-III	WP_045738409	*Xanthomonas* sp. MUS 060
CH-III	WP_045766978	*Xanthomonas albilineans*
CH-III	WP_046345054	*Xanthomonas campestris*
CH-III	WP_046418508	*Xylella fastidiosa*
CH-III	WP_046659518	*Lysobacter capsici*
CH-III	WP_046962405	*Xanthomonas pisi*
CH-III	WP_047127604	*Xanthomonas arboricola*
CH-III	WP_047130065	*Xanthomonas arboricola*
CH-III	WP_047138122	*Luteimonas* sp. FCS-9
CH-III	WP_047325464	*Xanthomonas translucens*
CH-III	WP_048491112	*Xanthomonas* sp. NCPPB1128
CH-III	WP_051247216	*Thermomonas fusca*
CH-III	WP_051412461	*Pseudoxanthomonas* sp. J35
CH-III	WP_051414286	*Pseudoxanthomonas suwonensis*

CH-III	WP_052101465	*Lysobacter arseniciresistens*
CH-III	WP_052103906	*Lysobacter concretionis*
CH-III	WP_052108135	*Lysobacter daejeonensis*
CH-III	WP_052116265	*Lysobacter dokdonensis*
CH-III	WP_052338318	*Luteimonas huabeiensis*
CH-III	WP_052497458	*Lysobacter* sp. A03
CH-III	WP_052631908	*Pseudoxanthomonas suwonensis*
CH-III	WP_053046187	*Xanthomonas*
CH-III	WP_053502634	*Xanthomonas campestris*
CH-III	WP_053835568	*Xanthomonas translucens*
CH-III	WP_053838099	*Xanthomonas translucens*
CH-III	WP_053841480	*Xanthomonas translucens*
CH-III	WP_054592909	*Xanthomonas arboricola*
CH-III	WP_054668329	*Stenotrophomonas acidaminiphila*
CH-III	WP_055251386	*Xanthomonas* sp. Mitacek01
CH-III	WP_055767150	*Stenotrophomonas*
CH-III	WP_055821746	*Xanthomonas* sp. Leaf131
CH-III	WP_055846990	*Xanthomonas* sp. Leaf148
CH-III	WP_055908485	*Lysobacter*
CH-III	WP_056110513	*Lysobacter* sp. Root690
CH-III	WP_056127581	*Lysobacter* sp. Root494
CH-III	WP_056177782	*Lysobacter*
CH-III	WP_056930973	*Stenotrophomonas acidaminiphila*
CH-III	WP_057505562	*Stenotrophomonas maltophilia*
CH-III	WP_057626930	*Stenotrophomonas terrae*
CH-III	WP_057634375	*Stenotrophomonas humi*
CH-III	WP_057638436	*Stenotrophomonas ginsengisoli*
CH-III	WP_057640071	*Stenotrophomonas daejeonensis*
CH-III	WP_057682407	*Xylella fastidiosa*
CH-III	WP_057919840	*Lysobacter antibioticus*
CH-III	WP_057944913	*Lysobacter gummosus*
CH-III	WP_057949459	*Lysobacter enzymogenes*
CH-III	WP_058564603	*Xylella fastidiosa*
CH-III	WP_058835845	*Luteimonas abyssi*
CH-III	WP_060747896	*Xanthomonas translucens*
CH-III	WP_062170560	*Stenotrophomonas*
CH-III	WP_064541099	*Xanthomonas translucens*
CH-III	WP_071869813	*Xylella fastidiosa*
CH-III	WP_074038914	*Xanthomonas*
CH-III	WP_074052858	*Xanthomonas vesicatoria*
CH-III	WP_074056420	*Xanthomonas gardneri*
CH-III	WP_074059552	*Xanthomonas vesicatoria*
CH-III	WP_074871521	*Lysobacter enzymogenes*
CH-III	WP_075245422	*Xanthomonas oryzae*
CH-III	WP_075250547	*Xanthomonas oryzae*
CH-III	WP_075272358	*Xanthomonas campestris*
CH-III	WP_075286631	*Xanthomonas campestris*
CH-III	WP_076037691	*Xanthomonas campestris*

CH-III	WP_076057531	*Xanthomonas campestris*
CH-III	WP_076586479	*Luteimonas tolerans*
CH-III	WP_078514311	*Xanthomonas*
CH-III	WP_078516911	*Xanthomonas*
CH-III	WP_078565160	*Xanthomonas axonopodis*
CH-III	WP_078566820	*Xanthomonas*
CH-III	WP_078589641	*Pseudomonas cissicola*
CH-III	WP_078759122	*Lysobacter spongiicola*
CH-III	WP_078997189	*Lysobacter enzymogenes*
CH-III	WP_080566033	*Xanthomonas campestris*
CH-III	WP_080576067	*Xanthomonas phaseoli*
CH-III	WP_080640308	*Xanthomonas campestris*
CH-III	WP_080715276	*Xanthomonas campestris*
CH-III	WP_080720725	*Xanthomonas*
CH-III	WP_080721482	*Xanthomonas euvesicatoria*
CH-III	WP_080870266	*Xanthomonas campestris*
CH-III	WP_080953598	*Xanthomonas phaseoli*
CH-III	WP_081023069	*Xanthomonas phaseoli*
CH-III	WP_081029607	*Xanthomonas*
CH-III	WP_081306912	*Xanthomonas campestris*
CH-III	WP_081340970	*Xanthomonas campestris*
CH-III	WP_081419617	*Xanthomonas cannabis*
CH-III	WP_081459345	*Pseudoxanthomonas suwonensis*
CH-III	WP_081711320	*Xanthomonas alfalfae*
CH-III	WP_082324397	*Xanthomonas oryzae*
CH-III	WP_082332432	*Xanthomonas citri group*
CH-III	WP_082861801	*Xanthomonas floridensis*
CH-III	WP_083992887	*Xanthomonas bromi*
CH-III	WP_084624954	*Xanthomonas cassavae*
CH-III	WP_086019466	*Xanthomonas vasicola*
CH-III	WP_086019800	*Xanthomonas*
CH-III	WP_086020489	*Xanthomonas vasicola*
CH-III	WP_087770686	*Xanthomonas oryzae*
CH-III	WP_087942545	*Xanthomonas oryzae*
CH-III	WP_087943472	*Xanthomonas sacchari*
CH-III	WP_087946473	*Xanthomonas fragariae*
CH-III	WP_087960510	*Xanthomonas translucens*
CH-III	WP_087960572	*Xanthomonas translucens*
CH-III	WP_087960648	*Lysobacter antibioticus*
CH-III	WP_087962321	*Xanthomonas arboricola*
CH-III	WP_088062373	*Xanthomonas fragariae*
CH-III	WP_088372132	*Xylella fastidiosa*
CH-III	WP_088578405	*Xylella fastidiosa*
CH-III	WP_089113455	*Xanthomonas citri*
CH-III	WP_091569116	*Oryzisolibacter propanilivorax*
CH-III	WP_091637744	*Lysobacter* sp. cf310
CH-III	WP_091806000	*Lysobacter* sp. yr284
CH-III	WP_093143971	*Pseudoxanthomonas* sp. GM95

CH-III	WP_093289887	*Pseudoxanthomonas* sp. CF125
CH-III	WP_095209323	*Luteimonas* sp. JM171
CH-III	WP_095522219	*Xanthomonas nasturtii*
CH-III	WP_095575360	*Xanthomonas hortorum*
CH-III	WP_096009818	*Xanthomonadaceae* bacterium NML93-0399
CH-III	WP_096207908	unclassified *Xanthomonadaceae (*miscellaneous*)*
CH-III	WP_096300262	*Luteimonas* sp. 100111
CH-III	WP_096418651	*Lysobacter capsici*
CH-III	WP_099802869	*Xanthomonas citri*
CH-IV	WP_057659948	*Pseudoxanthomonas dokdonensis*
CH-IV	AOA74089	*Stenotrophomonas rhizophila*
CH-IV	EWS79331	*Xylella taiwanensis*
CH-IV	OEZ01641	*Stenotrophomonas* sp. BIIR7
CH-IV	OHE89837	*Xanthomonadales* bacterium RIFOXYA1_FULL_68_6
CH-IV	SBV37254	uncultured *Stenotrophomonas* sp.
CH-IV	WP_019185500	*Stenotrophomonas maltophilia*
CH-IV	WP_038269899	*Xylella taiwanensis*
CH-IV	WP_038685524	*Stenotrophomonas rhizophila*
CH-IV	WP_046272604	*Stenotrophomonas maltophilia*
CH-IV	WP_055939244	*Pseudoxanthomonas* sp. Root65
CH-IV	WP_056881608	*Pseudoxanthomonas* sp. Root630
CH-IV	WP_057645443	*Stenotrophomonas panacihumi*
CH-IV	WP_057666977	*Stenotrophomonas koreensis*
CH-IV	WP_062351486	*Pseudoxanthomonas mexicana*
CH-IV	WP_066098294	*Xanthomonas massiliensis*
CH-IV	WP_068854887	*Stenotrophomonas rhizophila*
CH-IV	WP_070426213	*Stenotrophomonas* sp. LM091
CH-IV	WP_070470009	*Stenotrophomonas* sp. HMSC10F06
CH-IV	WP_079724481	*Pseudoxanthomonas indica*
CH-IV	WP_084738886	*Stenotrophomonas* sp. BIIR7
CH-IV	WP_088433832	*Stenotrophomonas maltophilia*
CH-IV	WP_093304156	*Pseudoxanthomonas* sp. CF385
CH-IV	WP_093489767	*Pseudoxanthomonas* sp. YR558
CH-IV	WP_093532760	*Stenotrophomonas rhizophila*
CH-IV	WP_095363320	*Stenotrophomonas rhizophila*
CH-IV	WP_097122168	*Pseudoxanthomonas wuyuanensis*

References

1 L. Fu, B. Niu, Z. Zhu, S. Wu and W. Li, *Bioinformatics*, 2012, **28**, 3150–3152.

2 W. Li and A. Godzik, *Bioinformatics*, 2006, **22**, 1658–1659.

3 A. Dereeper, V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J.-F. Dufayard, S. Guindon, V. Lefort, M. Lescot, J.-M. Claverie and O. Gascuel, *Nucleic Acids Res.*, 2008, **36**, W465–W469.

4 A. Dereeper, S. Audic, J.-M. Claverie and G. Blanc, *BMC Evol. Biol.*, 2010, **10**, 8.

5 R. C. Edgar, *Nucleic Acids Res.*, 2004, **32**, 1792–1797.

6 J. Castresana, *Mol. Biol. Evol.*, 2000, **17**, 540–552.

7 S. Guindon and O. Gascuel, *Syst. Biol.*, 2003, **52**, 696–704.

8 M. Anisimova and O. Gascuel, *Syst. Biol.*, 2006, **55**, 539–552.

9 F. Chevenet, C. Brun, A.-L. Bañuls, B. Jacq and R. Christen, *BMC Bioinformatics*, 2006, **7**, 439.

10 S. B. Needleman and C. D. Wunsch, *J. Mol. Biol.*, 1970, **48**, 443–453.

11 P. Rice, I. Longden and A. Bleasby, *Trends Genet.*, 2000, **16**, 276–277.

12 P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, *Genome Res.*, 2003, **13**, 2498–2504.

13 H. Kamisetty, S. Ovchinnikov and D. Baker, *Proc. Natl. Acad. Sci. U.S.A.*, 2013, **110**, 15674.

14 S. H. Ackerman, E. R. Tillier and D. L. Gatti, *PLOS ONE*, 2012, **7**, e47108.

15 L. Betancor, M.-J. Fernández, K. J. Weissman and P. F. Leadlay, *ChemBioChem*, 2008, **9**, 2962–2966.

16 F. Hubrich, M. Müller and J. N. Andexer, *J. Biotechnol.*, 2014, **191**, 93–98.

17 J. N. Andexer, S. G. Kendrew, M. Nur-e-Alam, O. Lazos, T. A. Foster, A.-S. Zimmermann, T. D. Warneck, D. Suthar, N. J. Coates, F. E. Koehn, J. S. Skotnicki, G. T. Carter, M. A. Gregory, C. J. Martin, S. J. Moss, P. F. Leadlay and B. Wilkinson, *Proc. Nat. Acad. Sci. U.S.A.*, 2011, **108**, 4776–4781.

18 F. Hubrich, P. Juneja, M. Müller, K. Diederichs, W. Welte and J. N. Andexer, *J. Am. Chem. Soc.*, 2015, **137**, 11032–11037.