ESM-Ezy: A deep learning strategy for the mining of novel multicopper oxidases with superior properties

Hui Qian[1,2,#], Yuxuan Wang[1,2#], Xibin Zhou[1#], Tao Gu[1,2], Hui Wang[1], Hao Lyu[1], Zhikai Li[1], Xiuxu Li[1,2], Huan Zhou[3], Chengchen Guo[1], Fajie Yuan[1,*] and Yajie Wang[1,2,3,4,5*]

[1]School of Engineering, Westlake University, Hangzhou 310014, Zhejiang, China

[2]The Center for Synthetic Biology and Integrated Bioengineering, Westlake University, Hangzhou 310014, Zhejiang, China

[3]Westlake Laboratory of Life Sciences and Biomedicine, Xihu District, Hangzhou 310024, Zhejiang Province, China

[4]School of life Science, Westlake University, Hangzhou 310014, Zhejiang, China

[5]Muyuan laboratory, Zhengzhou, Henan , China

[#]These authors contributed equally

[*]Corresponding authors: wangyajie@westlake.edu.cn, yuanfajie@westlake.edu.cn

Table of contents

Supplementary Tables

Supplementary Figures

Development of a deep learning model to identify MCOs with low sequence similarities

Table S1 The plasmids and strains used in this study

Strains and plasmid	Description
E. coli Top10	plasmid construction
E. coli BL21(DE3)	the host for protein expression
pET28a	the vector for protein expression
pET28a-sumo	the vector for protein expression

Table S2 All multicopper oxidases used in this study

Name	Entry	Organism
Eclac series		
Eclac(query)	P36649	*Escherichia coli* K-12
Sulfur	A0A2K2VPI4	*Sulfurimonas sp*
Salvi	A0A2N9X7P6	*Snodgrassella alvi*
Ppht	A0A238JCR0	*Pelagimonas phthalicica*
Ypes	Q8ZBK0	*Yersinia pestis biovar* Microtus str. 91001
Psoli	A0A328AKG1	*Phenylobacterium soli*
Faest	I0KHI7	*Fibrella aestuarina*
Mint	A0A4R6VBE7	*Mesocricetibacter intestinalis*
HR03 series		
HR03(query)	ACM46021	*Bacillus sp.* HR03
Aory	A0A5C7ESL7	*Azospira oryzae*
Talbi	A0A5B9EN11	*Terriglobus albidus*
Bfre	VEF49110	*Bacillus freudenreichii*
Tocean	D9RY72	*Thermosediminibacter oceani*
BMS3	A0A2H6K631	bacterium BMS3Abin01
Acidi	A0A0S8GAJ7	*Acidithiobacillales* bacterium SM23_46
DSM13 series		
DSM13(query)	Q65MU7	*Bacillus licheniformis* DSM13
Bcece	M7PAR1	*Bhargavaea cecembensis* DSE10
Scla	D5SLU3	*Streptomyces clavuligerus*
Arose	A0A7W9W9Q1	*Armatimonas rosea*
Cmeth	PWB55314	*Candidatus Methanoperedenaceae archaeon*
Cohnella	A0A433RH29	*Cohnella sp.* AR92

Table S3 The sequences of all multicopper oxidases mined in this study

Sulfur
ggatccTGTGAAAAGGGTAGTGATAAAAGCGTTGAACTGGATAAAAAAAGTGATGAAAC CCAGAAAGCCGCCCAGGAACCGCCGAAAAATATTGCCAATAATAAATTTGAGCCTTTCACC CAGAAACTGAAAATTCCGAAAGAAATTGATTTCGAACACGTTGCAAAAGCAAAATTTAATG CCCAGAAAAGCCTGAGCGCACTGTATAAAGAAAAAAAAACCGATATCCTGACCTTTCAGG GCGATCTGCCGAATCCGACCATTCGCATTAAAAATGGCGATGATTTTGAACTGGATTTTACC AATAGCCTGGAAAAACCGACCATTATTCATTGGCATGGCCTGCTGGTGCCGGAAGCAATGG ATGGCCATCCGAAAGATGCCATTGCCACCCAGATGCTGAAAGAATATCGCTATAAAGTGAAT CAGCGTGCCGGTACCTTTTGGTATCATACCCATCCGCATGGCCGCACCGGTGAAGAAATTTA TTATGGCCTGGCCGGTCTGTATATTATTGAAGATGATAATGAGAAGGCGCTGAATCTGCCGA GCGGCGAATTTGAACTGCCGCTGATTATTCAGGATCGTCGTTTTGATAAAGAAGGCGATCTG ATTTATAAGGAAACCCCGCAGGATAATAATGGCGTTCTGGGTGATGTTGTTATGGTGAATAG TACCGTGCATCCGTATAAAAATGTGAAAAATACCAAGTACCGCCTGCGCATTCTGAATGGTA GTAGCGCCCGTACCTATAAACTGGCATTTGAAGGTATTGAAGATTTTATGCTGATCGGCACC GATGGCGGTCTGCTGGAAGAACCGATTATTGTTAAAGATATTCTGATCGCAGTGGCCGAAC GTATTGATATTATTGTGGATTTTAAGGACAAGAAGGTGGGTGAAAGTGTTACCCTGAAAACC CTGGGTTTTAAAGAAGCAAATAATTTTGTGACCAACCCGGCATATCCGGATAGCGGTGCAA AAATGGATATTATGCGTTTTAAAGTGACCGAACTGAGTACCCAGAATAGTCAGATTCCGAAA AAACTGAGTACCATTGCCAAAATGAAAGCCAGTGATGCAAGTAAAAGCCGTACCATTACCA TGGAAATTATTGAAGGTGGCGTTTGGACCCTGAATAAAAAACCGTATGATATGCATCGTGTT GATGAAAAAGTTAAACTGGGTAGTACCGAAATTTGGGAAATTAAAAATAGCGCACATATGG CACATCCGTTTCATATGCATGGCGTTCATTTTCAGGTTCTGGAACGTACCAGCAGCATTGATT TTCCGACCGATAAAGGTTGGAAAGATACCGTTCTGGTTATGCCGCTGGAAAGTGTTCGTATT ATTGTTAAGTTTACCATCCCGGGTCTGTTTGTTCATCATTGCCATATTCTGGAACATGAAGAT CATAGTATGATGGCCAATTTTCTGGTTGAATAAaagctt
GSCEKGSDKSVELDKKSDETQKAAQEPPKNIANNKFEPFTQKLKIPKEIDFEHVAKAKFN AQKSLSALYKEKKTDILTFQGDLPNPTIRIKNGDDFELDFTNSLEKPTIIHWHGLLVPEAMDGHP KDAIATQMLKEYRYKVNQRAGTFWYHTHPHGRTGEEIYYGLAGLYIIEDDNEKALNLPSGEFE LPLIIQDRRFDKEGDLIYKETPQDNNGVLGDVVMVNSTVHPYKNVKNTKYRLRILNGSSARTY KLAFEGIEDFMLIGTDGGLLEEPIIVKDILIAVAERIDIIVDFKDKKVGESVTLKTLGFKEANNFV TNPAYPDSGAKMDIMRFKVTELSTQNSQIPKKLSTIAKMKASDASKSRTITMEIIEGGVWTLNK KPYDMHRVDEKVKLGSTEIWEIKNSAHMAHPFHMHGVHFQVLERTSSIDFPTDKGWKDTVLV MPLESVRIIVKFTIPGLFVHHCHILEHEDHSMMANFLVE
Salvi
ggatccCTGAATACCCTGAATCTGCCGGTGCTGGCCATTCCGCCGCTGCTGGATACCGGTA AAGATAATCGTATTCAGCTGACCATTCAGCAGGGTACCAGCCGTTTTGGCAATTATCAGACC CAGACCTGGGGTTATAATGGTGCCCTGCTGGGCCCGGCAATTCGTATTTTTCGCGGTAAACC GGTTCAGATTAAACTGCAGAATCAGCTGACCGAAGCCACCACCGTGCATTGGCATGGCCTG GAAGTTCCGGGCGAAGTGGATGGCGGTCCGCAGGCATGCATTCCGGCCGGTCAGGCCCGC GTTATTGAATTTACCAGCAGCCAGCCGGCAGCAACCTGCTGGTTTCATCCGCATCAGCATGG TCAGACCGGCCGCCAGGTTGCAATGGGTCTGGCCGGTCTGCTGCTGCTGGAAGATACCGTG AGTACCGGTCTGCCGATTCCGAAAAATTGGGGCCTGGATGATATTCCGCTGATTGTTCAGGA TAAAAAATTTAATACCGCGGGTCAGATTGATTATCAGCTGGATGTTATGAGCGCCGCAATTG

GTTGGTTTGGTGATACCCTGCTGTGCAATGGCCAGATTTATCCGCAGCATAATGCCCCGCGC
GGCTGGCTGCGTCTGCGTTTACTGAATGGCTGTAATGCCCGCACCCTGAATTTTGCATGTAG
CGATGGCCGTCCGCTGTATGTGATTGCCAGTGATGGTGGCTTTCTGCCGCAGCCGGTTACCG
TTAGTGGCCTGAGCGTGCTGCCGGGCGAACGTTTTGAAGTTATGATTGATACCAGTAATGGT
AAAGCATTTGATCTGCTGACCCTGCCGGTTAAACAGATGGGCATGAATCTGCCGCCGTTTG
ATCAGCCGTGTCCGATTGTGCGCGTGAGTGTTCTGAGCAGCCGCAATGATGCAAGCCTGCC
GGATCAGCTGACCAGCAGTCTGCCGGCACTGCCGGCATTAAGTCAGCTGGTTCATCGTCAT
TTTCAGCTGAGTATGGACCCTGAACTGGATAGCATGGGTATGCAGGCCCTGCAGCAGCTGG
CAGGCAAAAGCGCAATGCCGATGGAACATATGAGCATGCATCAGAGCCATATGGAAATGAT
GCATAATATGCATACCGATAAAAACGCCCTGGATCTGATGCATGCAAATCGTATTAATGGCCT
GACCTTTGATATGAATAAACCGGCCTTTGCAGCAAGCAAAGGCAAATATGAATGCTGGGAA
ATTAGTGGCCTGGGCGATATGATGCTGCATCCGTTTCATATTCATGGCACCCAGTTTCGTATT
CTGAATGAAAATGGTCAGCCGCCGGCAGCCCATCGCAGCGGTTGGAAAGATACCGTTCGC
GTGGAAGGTGGTGTGAGTAAAGTGCTGGTGAAATTTAATCATAGCGCCCCGGCCAGCTTTC
CGTATATGGCACATTGTCATATTCTGGAACATGAAGATACCGGTATGATGCTGGGCTTTACCG
TGTAAaagctt

LNTLNLPVLAIPPLLDTGKDNRIQLTIQQGTSRFGNYQTQTWGYNGALLGPAIRIFRGKPV
QIKLQNQLTEATTVHWHGLEVPGEVDGGPQACIPAGQARVIEFTSSQPAATCWFHPHQHGQTG
RQVAMGLAGLLLLEDTVSTGLPIPKNWGLDDIPLIVQDKKFNTAGQIDYQLDVMSAAIGWFGD
TLLCNGQIYPQHNAPRGWLRLRLLNGCNARTLNFACSDGRPLYVIASDGGFLPQPVTVSGLSV
LPGERFEVMIDTSNGKAFDLLTLPVKQMGMNLPPFDQPCPIVRVSVLSSRNDASLPDQLTSSLP
ALPALSQLVHRHFQLSMDPELDSMGMQALQQLAGKSAMPMEHMSMHQSHMEMMHNMHTD
KNALDLMHANRINGLTFDMNKPAFAASKGKYECWEISGLGDMMLHPFHIHGTQFRILNENGQ
PPAAHRSGWKDTVRVEGGVSKVLVKFNHSAPASFPYMAHCHILEHEDTGMMLGFTV

Ppht

ggatccCAGAGTGCCCCGCGTCCTCTGCCGCTGATTCCGCTGACCGATCTGACCGGTGGT
ATTGAAGGTCGTATTAGTCTGGCCCTGAATAAAGCCAGCCATGATTTTGGCACCGGTGCCCA
GAGTGAAACCTTTGGTATTAATCATGATTACCTGGGTCCGGTTCTGCGTGTGAAACAGGGTC
AGACCCTGCCGTTTGATGTTGTTAATAATATTGGCGAAACCAGCACCCTGCATTGGCATGGT
CTGCATATTCCGGGCGATGTTGATGGTGGCCCGCATCAGGAAATTGCACATGGCGCAACCT
GGAGTCCGGATGTGCCGATTGTTCAGCATGCAAGTATGAATTGGTTTCATAGTCATATGCAT
GGCCGCACCGCACGCCAGACCTATAGTGGCCTGGCCGGCGTTCTGCTGGTTGAAGATGATG
CAAGTCTGAGCGCAGATCTGCCGAAAACCTATGGTGTTGATGATTTTACCCTGATTCTGCAG
GATAAAATGTTTGATGGCAGTGGCAAAATGACCTATGCACTGACCGCCGAAGTGTTTGAAG
ATGGCTTTGAAGGTGATACCCTGACCGTTAATGGTGCAATTGCACCGGTTGCACAGAGTGT
GCCGACCGGCCTGGTTCGCCTGCGTATTCTGAATGCCTGTAATGCACGCTTTCTGGAACTGA
GCATGCAGAGCGGCCCGATTACCGTGATTGCAAGTGATGGCGGCTTTCTGGCAGCCCCGGT
GGAAGCCGCCAGCATTCTGATGAGTCCGGGCGAACGTTATGAAGTGCTGGTGGATATGAGC
ACCGTGGATAGCAATGAACTGACCGTGAATATGGATGGTGGCGGCAGCTTTTTTGCCAGTC
TGTTTGGTGGTAATGCAGCAACCACCGTGCTGACCCTGACCCGTCGCGCCGAAGCCGGTTT
TGATGGCGCAATGCCGCGCCGTCTGGCAAATCTGGCACCGCCGAAAGCCAGTGAAGCCAG
TGTTACCCGCAGTTTTCAGCTGGAAATGGATGTGGGCGCCGATCTGGCCGCCCTGGCAGCA
TCATGGGATAATTTTTGTGGTGATGCAGGTGCAATGGCCATTAATGGCCAGCCGATGAAAAT
GGATCGTATTGATGAAGTGGTGCGTAAAGGTAATACCGAAATTTGGCGCGTGAGCGTGGAT

GATCAGCTGCATCCGTTTCATATTCATGGCTGCAGCTTTCGTATTCTGAGCCAGAATGGCGC
AGCACCGCCGGCATATGCCGCCGGTTGGAAAGATATGGTTCATGTGGAAGAAGGCTGGAGC
GAAGTGCTGGTTCGCTTTGATTATGAAGCACCGAAAGATGCACCGTATATGTATCATTGCCA
TATTCTGGAACATGAAGATTGCGGTATGATGGGTCAGTTTACCGTTGGTTAAaagctt

QSAPRPLPLIPLTDLTGGIEGRISLALNKASHDFGTGAQSETFGINHDYLGPVLRVKQGQTL
PFDVVNNIGETSTLHWHGLHIPGDVDGGPHQEIAHGATWSPDVPIVQHASMNWFHSHMHGRT
ARQTYSGLAGVLLVEDDASLSADLPKTYGVDDFTLILQDKMFDGSGKMTYALTAEVFEDGFE
GDTLTVNGAIAPVAQSVPTGLVRLRILNACNARFLELSMQSGPITVIASDGGFLAAPVEAASILM
SPGERYEVLVDMSTVDSNELTVNMDGGGSFFASLFGGNAATTVLTLTRRAEAGFDGAMPRRL
ANLAPPKASEASVTRSFQLEMDVGADLAALAASWDNFCGDAGAMAINGQPMKMDRIDEVV
RKGNTEIWRVSVDDQLHPFHIHGCSFRILSQNGAAPPAYAAGWKDMVHVEEGWSEVLVRFDY
EAPKDAPYMYHCHILEHEDCGMMGQFTVG

Ypes

ggatccGCAGATTTTAGTCCGCTGCCGATTCCGCCGCTGCTGCAGCCTGATGCAAATGGTA
AAATTAATCTGAATATCCAGACCGGTAGCGTGGTGTGGCTGCCGAGCACCGCAACCCAGAC
CTGGGGTTATAATGGTAATCTGCTGGGCCCGGCAATTCGCCTGCAGCGTGGCAAAGCAGTG
ACCATTGATATTACCAATGCCCTGCCGGAAGCAACCACCGTTCATTGGCATGGCCTGGAAAT
TCCGGGTGAAGTTGATGGCGGTCCGCAGGCCCTGATTCAGCCGGGTGCAAAACGTCAGGT
TACCTTTGCAGTTGAACAGCCGGCAGCCACCTGCTGGTTTCATCCGCATACCCATAGTAAAA
CCGGCCATCAGGTTGCAATGGGCCTGGGTGGCCTGGTTCTGATTGATGATAGTGATAGCGA
AACCCTGCCGCTGCCGAAACAGTGGGGCGTGGATGATATTCCGGTTATTCTGCAGGATAAA
CTGCTGGATAAACATGGTCAGGTTGATTATCAGCTGGATGTGATGACCGCAGCAGTTGGCTG
GTTTGGTGATCGTATGCTGACCAATGGCGTTCCGTATCCGCAGCAGATTACCCCGCGTGGCT
GGGTGCGCCTGCGTTTACTGAATGGCTGTAATGCACGTAGTCTGAATCTGGCACTGAGTGAT
GGTCGTCCGATGTATGTTATTGCAAGTGATGGCGGTCTGCTGGCCGAACCGGTGGTTGTGC
GTGAACTGCCGATTCTGATGGGCGAACGTTTTGAAGTTCTGGTGGATACCCGCGATGGCCA
GAGCCTGGATCTGGTTACCCTGCCGGTGACCCAGATGGGCATGACCCTGGCCCCGTTTGAT
CAGCCGCTGCCGGTTCTGCGTATTCAGCCGAGTCTGGCAATTGGTAGCCAGGTTCTGCCGG
AAAGCCTGGTTGTTATTCCGGAACTGGCCGATGTGACCGGCGTTCAGGAACGCTGGTTTCA
GCTGATGATGGACCCTAAACTGGATATGCTGGGTATGCAGGCCCTGGTGGCACGCTATGGTA
TGAAAGCCATGGCAGGCATGAATATGAATCATGGTGATATGGGTGCCATGGATCATGGTAAT
CGTCCGGATATGAGTCAGGGCAAAATGAAAGGCATGGATCATGGCACCATGAATGGTGCCC
CGGCATTTAATTTTAGTCATGCAAATCGCATTAACGGTAAAGCATTTAGTATGACCGAACCG
GCATTTGATGCCAAACAGGGCAAATATGAAAAATGGACCATTAGTGGCGAAGGCGATATGA
TGCTGCATCCGTTTCATGTTCATGGTACCCAGTTTCGTATTCTGACCGAAAATGGTAAACCG
CCGGCCGAACATCGTCGCGGTTGGAAAGATATTGTTCGCGTTGAAGGCGCCCGCAGTGAA
ATTCTGGTGCGTTTTAATTATCTGGCACCGGCAAGTACCCCGTATATGGCACATTGCCATCTG
CTGGAACATGAAGATACCGGCATGATGCTGGGTTTTACCGTTAGCGCATAAaagctt

ADFSPLPIPPLLQPDANGKINLNIQTGSVVWLPSTATQTWGYNGNLLGPAIRLQRGKAVTI
DITNALPEATTVHWHGLEIPGEVDGGPQALIQPGAKRQVTFAVEQPAATCWFHPHTHSKTGHQ
VAMGLGGLVLIDDSDSETLPLPKQWGVDDIPVILQDKLLDKHGQVDYQLDVMTAAVGWFGD
RMLTNGVPYPQQITPRGWVRLRLLNGCNARSLNLALSDGRPMYVIASDGGLLAEPVVVRELPI
LMGERFEVLVDTRDGQSLDLVTLPVTQMGMTLAPFDQPLPVLRIQPSLAIGSQVLPESLVVIPEL
ADVTGVQERWFQLMMDPKLDMLGMQALVARYGMKAMAGMNMNHGDMGAMDHGNRPD

MSQGKMKGMDHGTMNGAPAFNFSHANRINGKAFSMTEPAFDAKQGKYEKWTISGEGDMML
HPFHVHGTQFRILTENGKPPAEHRRGWKDIVRVEGARSEILVRFNYLAPASTPYMAHCHLLEHE
DTGMMLGFTVSA

Psoli

ggatccCAACCTGGAAGGTTTACTACACCACTACCCGTTCCGCCACTGATCGATGCGCAA
GCACAGGGTAATAGCATTCGTTTGGTGGCGCAAACTGGTCGCCATGCGTTTGTTGAAGGTC
GTCCGGTTGCGACCTACGGCTACAGCGGTCCGGTTCTTGGTCCGACGATCCGTATGCAGTC
CGGTGCGCAAGTTGACGTTGCCATTGAGAACCGTATCGACACGGATACCATTGTACATTGG
CATGGTTTCCTGATTCCGTCCGAGCGCGACGGTGGTCCGCACGATCACATTGCTCCAGGTG
AAACCTGGCGTCGCGTGCTGCCGGTTCAGCAGCCGGAAACTACTGCGTGGTATCATCCGCA
CCCGCATCGTGACACCGGTCGTCAGGTTTATTTCGGCATGGCGGGTTTGGTGTTGATCGAG
GACGGTTCAGGCGCAAGATTAGACCTGCCACGTACCTATGGTGTCGACGACCTGCCTCTGG
TCCTGCAAGACCGTCTGTTTGATGGCAACGGTGATCTGATCTACCCGGCCAATCCGATGACC
GTAATGCAGGGCGCGCGTGGCAACACCATTATCGTGAACGGCGCGATCGCCCCCATCGCCC
GTGTTCCGGCGGGTCTCGTGCGTCTTCGATTGCTGAATGGCTCGAACGCTCGTAATCTCGAT
CTTGGCTTCGACGATGGCCGCCCGTTGCACGTGATTGCTAGCGACGGTGGCTACCTGGGCT
CCCCGGTGAGCATGACCCGTCTGCTGATCGCACCGAGCGAGCGCTTCGAGATCCTGGTCGA
CTTTAGCGATGGTAGGCCGGCGGTTCTGCAGACCGGCCCGGACCCGTTCCCGCCTATGATG
GGCATGATGATGGGGAGAGGCCAATCTGCTGCGGGTGATATTATGAGCTTCGCACCGGACC
GTGATCGCCCAGGTGCGAAGCGCAAAATCCCGAGCACCCTCGTGGACATGCCGGCGGCGG
CACCGAGTGCTCAACTGCTGCGTCGTCGTTTCGCCCTGAACGATATGGGCGGAATGATGGG
CGGTATGGGCGGCATGATGGGTCGTGGTATGATGGGCGGTATGATGGGCGGCGGCATGATG
GGTGGGGGCATGATGGGTGGCGGGATGATGGGTATGGGACGCGGCGGTTCCGGTCTGGGTA
TTAACGGCCAGTCTTTTGACATGCAACGTATCGATGTGGCTGTGGCGTTGGACAGCCGTGA
GGTTTGGGAGGTGTCGGCGCAGAGCATGGCGCACCCGTTCCACGTTCACGGCGTTCAATTT
CGTATTTTAAGCCTGGACGGTGCTGCTCCGCCTCCGCACCTGCAAGGTTGGAAGGATACGG
TGTTGGTGAGCCGCTCTGCGGAACTGCTCATCCACTTCACCCAGCCGGCGCTGCGCGCACA
CCCGTTTATGTTTCATTGTCACATCTTGGAACATGAAGATGCAGGTATGATGGGTCAGTATGT
TTGCGCATAAaagctt

QPGRFTTPLPVPPLIDAQAQGNSIRLVAQTGRHAFVEGRPVATYGYSGPVLGPTIRMQSGA
QVDVAIENRIDTDTIVHWHGFLIPSERDGGPHDHIAPGETWRRVLPVQQPETTAWYHPHPHRD
TGRQVYFGMAGLVLIEDGSGARLDLPRTYGVDDLPLVLQDRLFDGNGDLIYPANPMTVMQGA
RGNTIIVNGAIAPIARVPAGLVRLRLLNGSNARNLDLGFDDGRPLHVIASDGGYLGSPVSMTRL
LIAPSERFEILVDFSDGRPAVLQTGPDPFPPMMGMMMGRGQSAAGDIMSFAPDRDRPGAKRKI
PSTLVDMPAAAPSAQLLRRRFALNDMGGMMGGMGGMMGRGMMGGMMGGGMMGGGMM
GGGMMGMGRGGSGLGINGQSFDMQRIDVAVALDSREVWEVSAQSMAHPFHVHGVQFRILSL
DGAAPPPHLQGWKDTVLVSRSAELLIHFTQPALRAHPFMFHCHILEHEDAGMMGQYVCA

Faest

ggatccTGCAATAGTCATGATATGAGTGCAATGAATATGGGTGGCGAAACCGCCACCGCA
GTTACCGAAGGTAGTTTTACCACCCCGCTGCGCCTGCTGGATACCGCAAGCCCGAGTGGTC
CGCTGAGCGCCAAAAGCACCACCGAAGCAATTGTGGCAGGTAAAAATGCCCGTGTTCTGG
GTTATCGCGATGGTATGCTGGGCCCGACCTTTCGTGTTACCAGCGGCGCCACCGTTGATCTG
CGCTTTCTGAATGCCCTGACCGAAGAAACCAATATTCATTGGCATGGTCTGCTGGTTCCGGC
CAATATGGATGGCCATCCGGCACAGCTGGTGGGCGCCGGTCAGAGTTTTAATTATACCTTTC

GTCTGGATCAGGCAGCAACCATGGCATGGTATCATCCGCATCCGCATGGCAAAACCGCCAA
ACAGGCCTATATGGGTCTGGCCGGTCTGTTTATTGTTGAAACCCCGACCGAAAAAGCCCTG
AGCCTGCCGAGCGGTGCATATGAACTGCCGCTGATTCTGCAGGATAAACGTCTGGATGCCA
GCGGTAGCCCGCAGTATAATCCGAGTATGAATGATGTGATGCTGGGTTATATGGGCGAAATT
GTTACCGTTAATGGCGTGGCAAGTCCGATGCATAGTGTTGCCACCCGCATGTATCGCCTGCG
CCTGGTTAATGGCAGTAATGGCCGTCTGTATAATCTGGCCCTGAGTACCGGCGCACAGTTTT
GGGTGATTGGTAGTGATGGTGGCCTGCTGAGTGCCCCGGAAGGTGTTACCAGTCTGCTGCT
GGCCCCGGGCGAACGCGCTGATCTGCTGGTGGATTTTAGCAGCGTTCCGGTGGGTACCGAA
GTGTATCTGAAAAGCAATACCTTTAGCGGTGGTGGCGCACAGGGCCAGCAGGCATTTAATC
TGCTGAAATTTGTTGTGAGTAAGGCCGAAACCGAAACCTTTCGTTTACCGGCCCGTCTGGA
AACCGTGACCCCGCTGGCAGCAAGCGTTGCAACCAAAACCCGCACCTTTGATATTGGTATT
GCCATGCAGGCCATGCAGGGCATGAATATGACCGGCATGCATACCATTAATGGCAAAGTGTT
TAGCATGAATCGCATTGATGAAACCGTGAAACTGGGTGATACCGAAATTTGGGAATTTAATA
ATACCCAGGGCGATGAACCGCATCCGATGCATCTGCATGGTGCATTTTTTCAGGTGCTGAGC
CGCACCGGCGGTCGCAATGCACTGACCGCCAGTGAAAAAGGCTGGAAAGATACCGTTCTG
GTTATGCCGGGCGAACGTGTTCGTATTGTTGTGCCGTTTACCAAACCGGGCACCTTTGTGTT
TCATTGTCATAATCTGGAACATGGTGATGATGGCATGATGGGTAATTATCAGGTGGCATAAaag
ctt

CNSHDMSAMNMGGETATAVTEGSFTTPLRLLDTASPSGPLSAKSTTEAIVAGKNARVLGY
RDGMLGPTFRVTSGATVDLRFLNALTEETNIHWHGLLVPANMDGHPAQLVGAGQSFNYTFRL
DQAATMAWYHPHPHGKTAKQAYMGLAGLFIVETPTEKALSLPSGAYELPLILQDKRLDASGSP
QYNPSMNDVMLGYMGEIVTVNGVASPMHSVATRMYRLRLVNGSNGRLYNLALSTGAQFWVI
GSDGGLLSAPEGVTSLLLAPGERADLLVDFSSVPVGTEVYLKSNTFSGGGAQGQQAFNLLKFV
VSKAETETFRLPARLETVTPLAASVATKTRTFDIGIAMQAMQGMNMTGMHTINGKVFSMNRID
ETVKLGDTEIWEFNNTQGDEPHPMHLHGAFFQVLSRTGGRNALTASEKGWKDTVLVMPGERV
RIVVPFTKPGTFVFHCHNLEHGDDGMMGNYQVA

Mint

ggatccCGTGAACGCGCCCTGCTGCCTATTCCGCCGCTGCTGGAAGCCGATGCCGCAGGT
AATATTAATCTGAGTATTCAGGCAGGTCTGAGTGCATTTCTGGCAGGCAAACAGACCCCGA
CCTGGGGCTATAATGGCGCACTGCTGGGCCCGGCCCTGAAAGCTAAAGCAGGCCAGAATCT
GCGCGTGATGATTAATAATCGCCTGGATCAGCCGAGTACCCTGCATTGGCATGGTCTGGAAA
TTGGTGGTGAAGCAGATGGCGGTCCGCAGGCAGTGATTGCACCGCAGCAGCAGCGCGAAG
TGAAATTTCGCGTGGATCAGGCAGCAGCCACCTGTTGGTTTCATCCGCATACCCATCATCAG
ACCGGCTATCAGGTTGCAATGGGTCTGGGTGGCCTGTTTATTATTGAAGATGAATTTAGTCG
CGGTCTGAAACTGCCGGAACGTTGGGGTATTGATGATGTTCCGCTGATTCTGCAGGATAAA
CGTTTTGATCCGCAGGGTAATATTGATTATAAACTGGATGTGATGACCGCAGCAGTGGGTTG
GTTTGGCGATGTGATGCTGACCAATGGCAGCCTGTATCCGAAACATATTGCACCGAAAGGTT
GGCTGCGCCTGCGCCTGCTGAATGGTTGTAATGCCCGTAGCCTGCGTCTGGCAGCAAGTGA
TGGTCGTCCGATTTATGTTATTGCCGGCGATGGCGGCTTTCTGCCGGAACCGGTGGCCGTGC
AGGAACTGAGTCTGCTGATGGGCGAACGCTTTGAAGTTCTGCTGGATTGCAGCGATGGCAA
AGCATTTGATCTGCTGAGTCTGCCGGTTCATCAGATGGGTATGAATATGGCCCCGTTTGATC
AGCCGCTGGCACTGCTGAGCATTGAAACCGGTCTGATGCCGGGCACCGGCCGCTTACCGG
ATAGTCTGATTAGCCTGCCGCCGCTGCCGGTTACCGATAATCTGCCGGTGCGTCATCTGAAA
CTGGAAATGGATCAGAAACTGGATCATCAGGGTATGATGGCCCTGATGGCCCGTTATGGCCA

GCAGGCAATGCAGCAGGGTCATCATATGCATCATATGCACCATATGCAGGGCCAGGATGCCA
TGGCCACCGAAGAAGTGCAGATTATGGGTGCCAATCGTATTAATGGTCAGAGTTTTAGCATG
CATGAACCGATGTTTGATGTTAAAGTTGGCCAGTATGAAAAATGGATTATTAGCGGTACCGG
CGATATGATGCTGCATCCGTTTCATATTCATGGTACCCAGTTTCGCATTCTGACCGAAAATGG
CCAGACCCCGCTGCCGCATCGTCGTGGTCGCAAAGATATTGTGAAAGTGGAAGGCGGTGT
GAGCGAAGTTCTGGTGCAGTTTAAACATAAAGCAGATAAAGCAAACGCATATATGGCACAT
TGTCATCTGCTGGAACATGAAGATACCGGCATGATGATGAGTTTTACCGTTAGCAAATAAaag
ctt

RERALLPIPPLLEADAAGNINLSIQAGLSAFLAGKQTPTWGYNGALLGPALKAKAGQNLR
VMINNRLDQPSTLHWHGLEIGGEADGGPQAVIAPQQQREVKFRVDQAAATCWFHPHTHHQTG
YQVAMGLGGLFIIEDEFSRGLKLPERWGIDDVPLILQDKRFDPQGNIDYKLDVMTAAVGWFGD
VMLTNGSLYPKHIAPKGWLRLRLLNGCNARSLRLAASDGRPIYVIAGDGGFLPEPVAVQELSLL
MGERFEVLLDCSDGKAFDLLSLPVHQMGMNMAPFDQPLALLSIETGLMPGTGRLPDSLISLPP
LPVTDNLPVRHLKLEMDQKLDHQGMMALMARYGQQAMQQGHHMHHMHHMQGQDAMAT
EEVQIMGANRINGQSFSMHEPMFDVKVGQYEKWIISGTGDMMLHPFHIHGTQFRILTENGQTP
LPHRRGRKDIVKVEGGVSEVLVQFKHKADKANAYMAHCHLLEHEDTGMMMSFTVSK

HR03

ggatccATGACCCTGGAAAAATTCGTTGATGCCCTGCCGATTCCGGATACCCTGAAACCG
GTGCAGCAGAGTAAAGAAAAAACCTATTACGAGGTGACCATGGAAGAATGCACCCATCAG
CTGCATCGTGATCTGCCGCCGACCAGACTGTGGGGTTATAATGGTCTGTTTCCGGGTCCGAC
CATTGAAGTTAAACGCAATGAAAATGTGTACGTGAAGTGGATGAACAATCTGCCGAGTACC
CATTTTCTGCCGATTGATCATACCATTCATCATAGTGATAGCCAGCATGAAGAAAGTGAAGT
GAAAACCGTGGTGCATCTGCATGGCGGCGTTACCCCTGATGATAGTGATGGCTATCCGGAA
GCATGGTTTAGTAAAGATTTCGAACAGACCGGTCCGTATTTTAAACGTGAAGTGTATCATTA
CCCTAACCAGCAGCGCGGTGCAATTCTGTGGTATCATGATCATGCCATGGCACTGACCCGTC
TGAATGTTTATGCCGGCCTGGTGGGCGCCTATATTATTCATGATCCGAAAGAAAAGCGTCTG
AAACTGCCGAGTGATCGCTATGATGTGCCGCTGCTGATTACCGATCGCACCATTAATGAAGA
TGGCAGTCTGTTTTACCCGAGTGCACCGGAAAATCCGAGTCCGAGTCTGCCGAATCCGAGC
ATTGTTCCGGCCTTTTGTGGTGAAACCATTCTGGTTAATGGTAAAGTGTGGCCGTATCTGGA
AGTGGAACCGCGCAAATATCGCTTTCGTGTGATTAATGCCAGCAATACCCGCACCTATAATC
TGAGTCTGGATAATGGCGGCGAATTCATTCAGGTGGGTAGTGATGGCGGCCTGCTGCCTAG
AAGTGTGAAACTGAATAGCTTTAGTCTGGCACCGGCCGAACGCTATGATATTATTATTGATTT
CACCGCCTACGAGGGTGAAAGCATTATTCTGGCCAATAGCGCAGGTTGTGGCGGTGACGTT
AATCCGGAAACCGATGCCAATATTATGCAGTTTCGCGTTACCAAACCGCTGGCCCAGAAAG
ATGAAAGCCGCAAACCGGAATATCTGGCAAGTTATCCGAGCGTTCAGCATGAACGCATTCA
GAATATTCGTACCCTGAAACTGGCCGGTACCCAGGATGAATATGGCCGCCCTGTTCTGCTGC
TGAATAATAAGCGTTGGCATGATCCGGTGACCGGCGCACCTAAAGTTGGTACCACCGAAAT
TTGGAGCATTATTAATCCGACCCGTGGCACCCATCCGATTCATCTGCATCTGGTGAGTTTTCG
CGTGCTGGATCGTCGCCCGTTTGATATTGCACGCTATCAGGAAAGCGGCGAACTGAGTTATA
CCGGTCCGGCAGTGCCGCCTCCTCCTAGCGAGGAGGGTTGGAAAGATACCATTCAGGCAC
ATGCCGGCGAAGTGCTGCGCATTGCAGCTACATTTGGTCCGTATAGTGGCCGCTATGTTTGG
CATTGCCATATTCTGGAACATGAAGATTACGATATGATGCGCCCGATGGATATTACCGATCGT
CATAAATAAaagctt

MTLEKFVDALPIPDTLKPVQQSKEKTYYEVTMEECTHQLHRDLPPTRLWGYNGLFPGPTI

EVKRNENVYVKWMNNLPSTHFLPIDHTIHHSDSQHEESEVKTVVHLHGGVTPDDSDGYPEAW
FSKDFEQTGPYFKREVYHYPNQQRGAILWYHDHAMALTRLNVYAGLVGAYIIHDPKEKRLKL
PSDRYDVPLLITDRTINEDGSLFYPSAPENPSPSLPNPSIVPAFCGETILVNGKVWPYLEVEPRKY
RFRVINASNTRTYNLSLDNGGEFIQVGSDGGLLPRSVKLNSFSLAPAERYDIIIDFTAYEGESIILA
NSAGCGGDVNPETDANIMQFRVTKPLAQKDESRKPEYLASYPSVQHERIQNIRTLKLAGTQDE
YGRPVLLLNNKRWHDPVTGAPKVGTTEIWSIINPTRGTHPIHLHLVSFRVLDRRPFDIARYQES
GELSYTGPAVPPPPSEEGWKDTIQAHAGEVLRIAATFGPYSGRYVWHCHILEHEDYDMMRPM
DITDRHK

Aory

ggatccATGGCCCAGCCGCCTAAATGGACCCCGCCGGCTCATGCCCCGCTGGACCCTAGA
AGTCAGAAACAGTTTGTGAATCCGCTGCCGATTCCGCCGGTGGCACGCCCTGCTGTTGCAG
ATGGCACCCATTATGAAATTCCGATTGGCGAATTTCGCCAGCATCTGGGCATTCGCGATCCG
GAAACCGGCGCCCCGCTGATGACCACCGTGTGGGGTTATGCCGGTCGTTATCCGGGCCCGA
CCATTGAAGCACGTCGCGGTCGTCCGATTACCGTTCGCTGGGTGAATGCACTGTATGGTGCC
AGCGGTCCGCTGCCGCATCGCATGCCTGTTGATACCAGTGTTCATATGGCCCCGCTGCGTAA
TTGGCCGGCCAGCGGTGTTCCGACCGTTACCCATCTGCATGGTGGTCATACCGAATGGCAG
AGTGATGGTGATCCGGATGCATGGTATACCCCGGGCTATGCACAGAAAGGCCCGCTGTTTG
CAAAAGAAGTGTATACCTATGATAACGATCAGGAAGCCGCACTGCTGTGGTATCATGATCAT
GCACTGGGTTTTACCCGCCTGAATGTTTATACCGGCCTGGCCGGTCTGTATGTGATTCGTGA
TGATTGGGAACGCAGCCTGGGTCTGCCGAGTGGTCGTTATGAACTGCCGCTGCTGATTCAG
GATCGTAGTTTTTATGCCAATGGTGAACTGTATTATCATGCCCTGGCCCAGGAACCGAGTCA
TCCGAGTCCGAGTGTTCTGCCGGAACAGTTTGGTGATGTGATTCTGGTGAATGGTATGGCCT
GGCCGACCGCCGAAGTTGAACCGCGTAAATATCGCCTGCGTCTGCTGAATGGCAGTGATAG
TCGTTTTTATCGCCTGGGTTTTAGTAGTCGCCTGAAATTTATTCAGATTGGCAGTGATGGTGG
TCTGCTGGATGCCCCGGTTGAACTGACCGAACTGCTGCTGGCACCGGCCGAACGCGCCGAT
GTGATTGTGGATTTTGCCCCGCTGCGCGGTAAAACCGTTCTGCTGCTGAATGATGCCCGCGC
CCCGTTTCCGGATGGTGATCCTGTGGACCCTCGTACCAATGGTCGTGTTATGGCCTTTAAAG
TGGGCAGTACCACCACCAAAGATGAAAGTCGTATTCCGGATCGCCTGCGTCGCGAACCGAT
TCCGTTTCTGGTTCCGACCGCACCGAAACGTGCCCTGCTGCTGTGGGAAGGTCAGGATGCA
CATGGCCGCCAGAAAAGCCTGCTGGGCACCGTGCAGGATGGTGCACTGCATTGGAATAGTC
CGATTACCGAAAATCCGGCCCTGGGTGCAACCGAAGTGTGGGAATTTTATAATACCACCCC
GGATACCCATCCGGTGCATATTCATCTGGTTCATTTTCGTATTATGAGTCGTCAGAAATTTAA
GGCCGATCTGGATGAAAAAACCGGCCGCCTGAGCAATATTCGTTTTACCGGTCCGCCGCTG
CCGCCGCCTCCTGAGGAGCGTGGTTGGAAAGATACCGTTCGTGCAAATCCGGGCGAAGTTA
CCCGTGTGATTGCAACCTTTGATCGCCCGGGCCGTTATGTTTGGCATTGTCATATTCTGAGCC
ATGAAGATCATGAAATGATGCGTCCGTATCAGGTGGGCCCGATTGGTGAAACCCCGGGTCG
TGGTGCACTGTAAaagctt

MAQPPKWTPPAHAPLDPRSQKQFVNPLPIPPVARPAVADGTHYEIPIGEFRQHLGIRDPETG
APLMTTVWGYAGRYPGPTIEARRGRPITVRWVNALYGASGPLPHRMPVDTSVHMAPLRNWPA
SGVPTVTHLHGGHTEWQSDGDPDAWYTPGYAQKGPLFAKEVYTYDNDQEAALLWYHDHAL
GFTRLNVYTGLAGLYVIRDDWERSLGLPSGRYELPLLIQDRSFYANGELYYHALAQEPSHPSPS
VLPEQFGDVILVNGMAWPTAEVEPRKYRLRLLNGSDSRFYRLGFSSRLKFIQIGSDGGLLDAPV
ELTELLLAPAERADVIVDFAPLRGKTVLLLNDARAPFPDGDPVDPRTNGRVMAFKVGSTTTKD
ESRIPDRLRREPIPFLVPTAPKRALLLWEGQDAHGRQKSLLGTVQDGALHWNSPITENPALGAT

EVWEFYNTTPDTHPVHIHLVHFRIMSRQKFKADLDEKTGRLSNIRFTGPPLPPPPEERGWKDTV
RANPGEVTRVIATFDRPGRYVWHCHILSHEDHEMMRPYQVGPIGETPGRGAL

Talbi

ggatccATGACAGAATTTACTAGGCCCCTACACTCACAGTTGCCGCCTGCACGTCTGTGG
GGTTACGAGGGTCAGTATCCGGGTCCGATTATCGACACCCGCCGTAATAAACCGATCGCGGT
TCAGTGGGAGAACCGCCTGCCGACGCGTCATATATTACCGATCGACCCTCATATTCACGGTG
CTATGCCGCCGACCCCGGAAGTTCGCACTGTGCCGCACCTGCACGGAGCCAATGTTCCGTC
CCCGTCCGACGGCTTGCCGGAAAAGTGGTTCACCCCGGGTCATAGCGCACGTTATGAATAT
CCGAATCGCCAACGTGCGGCAACCCTGTGGTATCATGATCATGCAATCGGCATCACCCGTCT
TAACGTTTATGCCGGTTTGAGCAGCTTTTATCTGCTGAGGGACGACGAGGAGCTGGGTATG
CACTTGCCGTCTGGTGAGTACGAGATCCCATTGGTTCTGCAGGATCGTACCTTAGACGAGC
ACGGCCAACTCCTGTACGCCCCGACCCTGGACGACGCCGTACCGCTGCCGAAAGGTGTTT
GGGGTCCGATGTTTTTCGGTGAATTGCCGGTCGTCAATGGGGCGATTTATCCATACGTGGAA
GTGAGACCGCAGTTATACCGTATTCGTTTGCTCAACAGTTCTAACTCTCGCGTCCTGAACCT
GTACCTGAACCTGGCGAAATCCCCGACGGATATCCCGCAATTGATCAAATTCCACCAGATTG
GCAGCGACGGCGGTCTGATGCCGCGTCCGGTGGCGCTGGAGAAGCTGGAGCTGGCGCCTG
GCGAAAGAGCGGATCTGATTGTTGATTTCTCCAGCTTAGCTGGTAAGACCGTGACGCTGAG
CAACGATGCTCCGGCTCCGTACCCAGGCTGGAATGTGATGAATGCAACTTGGCCGATCCTG
CTGGAATTTATGCAGTTTCGTGTTACCCTCCCTGCTGAACCGTCGCGATCATTCAGCCTGCC
ACAAAACGTGACGTTTGCCAAGCTGGACGAGGGCGAGGCGATCCGTACCCGCGACTTCAT
CTTGACCGAAGAGATGGATGCGCAAGGTCGTAGCGTTGGCTTGCACATTAACGGCAAGGGT
TACCATGACCCGGTGACCGAAACCGTGACCCTTGGCACCCTCGAGAAGTGGCGTTTCATCA
ACACCAGCGACGATGCGCATCCGATGCATCTTCACCTGGTGCAATTCCAGATTCTGCACCGT
CAAGGTTTTAACCTGGGCACCTACCGCATGAACGGCAAGATCGAACCGGTGGGTTTGACGC
GTCAGCCGCGCGCGAATGAACAGGGTTGGAAAGATACCGCAACGGTAAACCCGGGTGACA
TTTTGACCATTCTGGCGCGCTTCGAAGGCTACACCGGTAAATACGTTTTTCACTGCCATATG
CTGGAGCACGAGGATAACGATATGATGCGTCCATTTGTTGTCGTGGCTCCGGGCCAAGGCT
AAaagctt

MTEFTRPLHSQLPPARLWGYEGQYPGPIIDTRRNKPIAVQWENRLPTRHILPIDPHIHGAMP
PTPEVRTVPHLHGANVPSPSDGLPEKWFTPGHSARYEYPNRQRAATLWYHDHAIGITRLNVYA
GLSSFYLLRDDEELGMHLPSGEYEIPLVLQDRTLDEHGQLLYAPTLDDAVPLPKGVWGPMFFG
ELPVVNGAIYPYVEVRPQLYRIRLLNSSNSRVLNLYLNLAKSPTDIPQLIKFHQIGSDGGLMPRP
VALEKLELAPGERADLIVDFSSLAGKTVTLSNDAPAPYPGWNVMNATWPILLEFMQFRVTLPA
EPSRSFSLPQNVTFAKLDEGEAIRTRDFILTEEMDAQGRSVGLHINGKGYHDPVTETVTLGTLE
KWRFINTSDDAHPMHLHLVQFQILHRQGFNLGTYRMNGKIEPVGLTRQPRANEQGWKDTATV
NPGDILTILARFEGYTGKYVFHCHMLEHEDNDMMRPFVVVAPGQG

Bfre

ggatccATGAAGATCAAGAACAAGCGTCGTATGAAAGTGGACCCTAGCAATCCGGAAAC
CATTCCGAAATATATGGATGAACTGCCGATTCCGCCGGTTGCACGTCCGCTGGCAGAAATTA
AAGGTAGCCCGTATTATGAAATTGCCATGCGTCAGGTGCCGCATCGCTTTCATCGTCTGTTT
CCGCCGACCACCGTTTGGGGCTATGATGGTATGCTGCCGGGTCCGACCATTAAAGTGCAGA
AAGATGAAAAAATCTACGTGCGTTGGAAAAATAAACTGCCGGAAAAACATCTGCTGCCGAT
TGATCGCACCCTGCATGAAACCGCAGGTCCGCCGGATGTGCGCACCGTGGTGCATCTGCAT
GGTGCAAATGTGGCATGGGATAGTGATGGTCATCCGGAAGCCTGGTTTAGTCGCGATTTTGC

CAAAACCGGTGCCACCTTTCGTCGCAAAGTTTATGAATATACCAATAAACAGATGGGTGCA
ACCCTGTGGTATCATGATCATGCCATTGGCATTACCCGCCTGAATGTGTATAGTGGTCTGAGC
GGCTTTTATCTGATTGAAGATCCGGTGGAAAAACATTTAAAACTGCCGAAAGATGGTTATGA
TATTCCGCTGATGATTCAGGATCGCAGCTTTCGTAGTGATGGTAGTCTGAGTTATCCGGAAA
ATACCAATCCGCCGGCCCCGGTGAATCCGAGCGTTCAGCCGTTTTTTATTGGTAATACCATT
GCAGTTAACGGTAAAATTTGGCCGAAACTGACCGTGGAACCGCGTAAATATCGCTTTCGTAT
TCTGAATGCAAGCAATACCAATGCCTATACCCTGCGCCTGGGTGATGGTCGCAAATTTTATC
AGATTAGCACCGATGGTGGCCTGCTGACCGAACCGGTTGAACTGACCACCCTGCCGCTGG
AACCGGCAGAACGCAGTGATGTGATTATTGATTTTAGTCAGCATAAGGGTAAGAAACTGAT
TCTGCAGAATACCAATGCGGAAGGCAATATGGGCATTATTATGCGCTTTGATGTTCTGCAGC
CGCTGCGCGGTCGCGATACCAGTGAAATTCCGGCAAAACTGATTAGCGAAGAACAGGTTCT
GTATGAACATCATGCAGATAAAACCCGTCTGCTGAAACTGGATGCAATTCAGGATGAATATA
ATCGCCCGGTTCTGCTGCTGGATGATCGCATGTGGCATGATCCGGTGACCGAAAAACCGGT
TATTGGTGATACCGAAGTGTGGAAACTGATTAATGTTACCAATTTTGCCCATCCGATTCATAT
TCATCTGATTCAGTTTAAAATCCTGCATCGCACCCCGTTTGATCTGGAACGCTTTCAGCAGG
ATGGCTATATTGATTATACCGGTCCGCCGATTGAACCGGCCGTTCATGAACGTGGTTGGAAA
GATACCGTGAAAGCCGAACCGGGTATGGTTACCAGCGTGATTATGAAATTTACCGAAAATCC
GGGTGAATATGTTTGGCATTGCCATATTCTGGAACATGAAGATTATGATATGATGCGCCCGAT
GCGCGTGGTGGAAAAAGAAAAATAAaagctt

MKIKNKRRMKVDPSNPETIPKYMDELPIPPVARPLAEIKGSPYYEIAMRQVPHRFHRLFPP
TTVWGYDGMLPGPTIKVQKDEKIYVRWKNKLPEKHLLPIDRTLHETAGPPDVRTVVHLHGAN
VAWDSDGHPEAWFSRDFAKTGATFRRKVYEYTNKQMGATLWYHDHAIGITRLNVYSGLSGFY
LIEDPVEKHLKLPKDGYDIPLMIQDRSFRSDGSLSYPENTNPPAPVNPSVQPFFIGNTIAVNGKIW
PKLTVEPRKYRFRILNASNTNAYTLRLGDGRKFYQISTDGGLLTEPVELTTLPLEPAERSDVIIDF
SQHKGKKLILQNTNAEGNMGIIMRFDVLQPLRGRDTSEIPAKLISEEQVLYEHHADKTRLLKLD
AIQDEYNRPVLLLDDRMWHDPVTEKPVIGDTEVWKLINVTNFAHPIHIHLIQFKILHRTPFDLE
RFQQDGYIDYTGPPIEPAVHERGWKDTVKAEPGMVTSVIMKFTENPGEYVWHCHILEHEDYD
MMRPMRVVEKEK

Tocean

ggatccATGCATACCATGAATTGTAAGGTGAATCAGAATTTTATCCCGTTTTATAAGGAGAT
GATCAAACTGGATCGTGCAATTCTGTATCTGTATGGTGGCCGTGGCTATTATATGGCCCCGAA
ACTGGAAAAATTTAAAGATCCGCTGCCGGTTCCGAAATTTATTCGCCCGGTGGGCACCAAA
AATGGCATTCCGTATTATGAACTGGAAATGACCGAATTTTACCAGAATCTGCATAGTGATATG
CCGGATACCCGCGTGTGGGGCTATGAAAATAGCTATCCGGGCCCGACCATTGAAGCATATAA
AGGCCAGCCGATTTATGTGCGTTGGATTAATCGCCTGCCGGAAAAACATTTTCTGCCGATTG
ATAAAACCATTCATGGTGCCAAAGATAATCCGGAAGTTCGCACCGTGGTGCATCTGCATGGT
CTGAATGTTCGTCCGGATAGCGATGGCTTTCCGGATGATTGGTTTACCCCGGGTAAAAGTGC
ACTGTATTATTATCCGAATAAGCAGCAGGCCGCAACCCTGTGGTATCATGATCATGCAGTTG
GTATTACCCGTCTGAATGTGTATGCAGGTCTGGTTGGTCTGTATCTGATTCGCGATGAACGT
GAAGAACGTCTGAATCTGCCGAGCGGCGAATATGAAATTCCGCTGATTATTCAGGATAAAG
ATTTTAATGACGACGGTAGCCTGAGTTATCCGGCACCGAATAATAATGGTGTGGAACCGAGT
ATTGTTCCGGGCTTTTTTGGTAATTATGCACTGGTGAATGGCAAAGTTTGGCCGTATCTGGTT
GTGAAACCGCGTAAATATCGCTTTCGCATTCTGAATGGCAGTAATAGCCGTAGCTATAAACT
GCGCTTTAGTGGTGAAGCCGGTCGTATTCTGCCGGCCTGGTATCAGATTGGCACCGATGGC

GGTTTTCTGGAACGTCCGGTGAAACTGGATAGCCTGAGCCTGCAGCCGGCCGAACGCGCT
GATGTTATTGTGGATTTTACCGGTCTGGAAAATCAGACCTTTACCCTGATTAATGAAGAAATT
CCGCCGTTTGGCCCGCCGCTGCCGGAAATTATGCAGTTTCGCGTGAGTGGCACCCCGGTGA
AAGATAATAGTAGCCTGCCGCTGATTCTGAATCGTATTAAACCGCTGCCGGTGAAAAAAGC
CAAACAGCGCAATATTGAAATTGTGGTTGGTCAGGATGAACTGGGCCGTTTTATGTTTATGC
TGGAAAATAAAAAGTGGACCGATCCGGTTACCATTAAAACCAAACTGGGCAATGTTGAAGT
GTGGAATATTATTAATACCGCCGCAGCCGCCCATCCGATTCATGTGCATCTGGTTCAGTTTCA
GATTCTGAATCGCCAGCCGTTTGATGTTCAGGAATATACCGCAACCGGCACCCTGAAATTTA
TTGGTCCGCCGGTGCTGCCGGATGAAAATGAAAAAGGCTGGAAAGATACCGTGCGTGCCG
AACCGGGCCATGTGACCCGCATTATTGCCCGTTTTGGCGATTTTACCGGCATTTATCCGTTTC
ATTGCCATATTCTGGAACATGAAGATCATGAAATGATGCGCCCGTTTGAAGTGTTTAAACAT
AAAAGCGGTGGTAAAGATAAGAAAGATAGTGAAGAAATCCACACCGTTAAACCGGATACC
AAAGCAGATAAAGAAGAACCGGCCAAATAAaagctt

GSMHTMNCKVNQNFIPFYKEMIKLDRAILYLYGGRGYYMAPKLEKFKDPLPVPKFIRPVG
TKNGIPYYELEMTEFYQNLHSDMPDTRVWGYENSYPGPTIEAYKGQPIYVRWINRLPEKHFLPI
DKTIHGAKDNPEVRTVVHLHGLNVRPDSDGFPDDWFTPGKSALYYYPNKQQAATLWYHDHA
VGITRLNVYAGLVGLYLIRDEREERLNLPSGEYEIPLIIQDKDFNDDGSLSYPAPNNNGVEPSIVP
GFFGNYALVNGKVWPYLVVKPRKYRFRILNGSNSRSYKLRFSGEAGRILPAWYQIGTDGGFLE
RPVKLDSLSLQPAERADVIVDFTGLENQTFTLINEEIPPFGPPLPEIMQFRVSGTPVKDNSSLPLIL
NRIKPLPVKKAKQRNIEIVVGQDELGRFMFMLENKKWTDPVTIKTKLGNVEVWNIINTAAAA
HPIHVHLVQFQILNRQPFDVQEYTATGTLKFIGPPVLPDENEKGWKDTVRAEPGHVTRIIARFG
DFTGIYPFHCHILEHEDHEMMRPFEVFKHKSGGKDKKDSEEIHTVKPDTKADKEEPAK

BMS3

ggatccTGCGGTGATAATCAGCAGACCGCCGCAACCACCCGCCCGGGTGATCCTGTGACC
ACCAGTAAAGTTATTCGTACCCCGGCACGTATTGATCATTTTATTGATCCGCTGCCGGTGCCG
CCGACCCTGAGTCCTGATACCGAACGCTATCCGGATACCGATTATTATGAAATTCGTATGAGT
CAGTTCGGTCAGGAACTGCATAGCCAGCTGCCGGAAACCACCGTTTGGGGCTATGAAAAA
AGTTTTCCGGGCCCGACCATTGAAGCCCGTCGTGGCCGTCGTACCCGCGTTAAATGGATTA
ATGATGATCTGCCGGCCGCCCATCTGCTGCAGGCCTCAATTGATCCGACCATTTATCAGGGC
CGTGAATTTCCGGATGTGCGCACCATTGTGCATCTGCATGGTAGCTTTGTTCGTCCGGAATT
TGATGGCCAGCCGGATGCCTGGAGCAGCCCGGGTGCAGGTAATACCGGCCCGGGTCATTTT
GGTAGTGAATTTACCTATCCGAATGATCAGCCGGCCACCATGCTGTGGTATCATGATCATGC
AATGCATAGCACCCGCCTGAATGTTTATGCCGGTCTGGCCGGTCTGTATTTTATTCGCGATGA
AGAAGAAGAAAAGCTGAGTCTGCCGGTTGGCAATCATGAAATTCCGCTGGTTATTCAGGAT
CGCACCCTGGCCGATGATGCAAGTCTGAGTTATCCGACCACCGGCATTACCCCGGTTCATCC
GGTGTGGGTGATGCGCTTTCTGGGTGATATGCCGCTGATTAATGGTAGCGCCTATCCGTATCT
GGATGCCGAACCGCGCCGTTATCGCCTGCGCCTGCTGAATGCCAGTAATACCCGCACCTGG
AATCTGTGGTTTGATGCCGGTGGTGGCCCGTTTCCGTTTCATGTGATTGGTAGTGATGGTGG
TTTTCTGCCGGCCCCGGCCCGCGTTGAAAGTCTGCGTCTGGCCCCGGCAGAACGCGCAGAT
GTGATTCTGGATCTGAGCGCAGCCGATCCGGGCACCGTTTTTACCCTGCGTAATGATGCCCC
GGCACCGTATCCGGAAGGTGGCGATCAGCCGCTGGATAATCTGATGCAGATTCGTCTGAGC
CGTGATCTGGATGGTGAAGATCGCAGTACCCCGGCAGATCGTCTGAGTCTGCCTGCAACCA
GCACCGCAACCCCGACCAGTGGCGTTCCGAAACGCATTTTTTATCTGGATGAACTGAGCAA
TGATCAGGGCAGCGCAGTGGAACTGCTGATTAATGATCGCCGTTTTCATGATCCGGTTGAA

GAACAGCCGGCAGCAGGCACCACCGAAATTTGGGAAATTGTGAATCTGACCCTGGATGCC
CATCCGTTTCATATTCATCTGGTGCAGTTTCAGGTGATGAATCGCCAGCCGCTGGACCAGGC
ACAGTATTGGAAAGATAAAGATGCATATGTTGCCGGCACCGCACCGATTCCGGACCCTCATA
CCTATCTGACCGGCGCAGCAGTGCCGCCGGCACCTGCAGAAACCGGTTGGAAAGATACCG
CCCTGAGCATGCCGGCCGAAGTTCTGCGTCTGGCGGTGCCGTTTAGCCTGCCGGCCGGTGT
GAGTGGCCCGGCAAGATATATTTATCATTGCCATATTCTGGAGCATGAAGATAATGATATGAT
GCGCCCGTTTGATGTTGTTTAAaagctt

CGDNQQTAATTRPGDPVTTSKVIRTPARIDHFIDPLPVPPTLSPDTERYPDTDYYEIRMSQF
GQELHSQLPETTVWGYEKSFPGPTIEARRGRRTRVKWINDDLPAAHLLQASIDPTIYQGREFPD
VRTIVHLHGSFVRPEFDGQPDAWSSPGAGNTGPGHFGSEFTYPNDQPATMLWYHDHAMHSTR
LNVYAGLAGLYFIRDEEEEKLSLPVGNHEIPLVIQDRTLADDASLSYPTTGITPVHPVWVMRFL
GDMPLINGSAYPYLDAEPRRYRLRLLNASNTRTWNLWFDAGGGPFPFHVIGSDGGFLPAPARV
ESLRLAPAERADVILDLSAADPGTVFTLRNDAPAPYPEGGDQPLDNLMQIRLSRDLDGEDRSTP
ADRLSLPATSTATPTSGVPKRIFYLDELSNDQGSAVELLINDRRFHDPVEEQPAAGTTEIWEIVNL
TLDAHPFHIHLVQFQVMNRQPLDQAQYWKDKDAYVAGTAPIPDPHTYLTGAAVPPAPAETGW
KDTALSMPAEVLRLAVPFSLPAGVSGPARYIYHCHILEHEDNDMMRPFDVV

Acidi

ggatccATGGATAACGTTCTGGTGCCGGATACCACCACCTATCCGGGCACCGATTATTATC
ATGTTCGCATGGTTCAGGTTCAGCAGGATCTGGGTCTGATTGATCCGGAAACCCGTAAACC
GCTGCGTACCACCGTTTGGGCATATGGCGCCGCCAATCGTACCGCAACCTATCCGGGTCCG
ACCATTGTTGCCCATAGTACCCTGGCAACCAATCGCCAGCCGGGCAAACCGGTGAAAGTGC
GCTGGGAAAATGCACTGCCGGATAAACATCTGCTGCCGGTTGATACCACCGTTCATTGTGG
CCCGGATGCACGCCAGCAGCATAGCCATTGTCGTCCGTTTGTGCGCACCGTTACCCATCTGC
ATGGCGGTCATGTGCCGGATCATAGTGATGGTTATCCGGAAGCATGGTTTAGCCCGGGTTTT
CGTGAAAAAGGTCCGCTGTGGAGTCGTGAAGTGTATGATTATCCGAATGATCAGGAAGCCG
CCACCCTGTGGTATCATGATCATGCCATGGGTATTACCCGTCTGAATGTGTATGCAGGCCTGG
CAGGTTTTTATCTGGTGCGTGATGATAATGAACAGCGTCTGCAGCGTGATGGTCGCCTGCCG
GCCCCTGCTTATGATGTGCCGCTGGTGATTCAGGATCGTAGTTTTCGTCGTGATGGCAGTCT
GTTTTATACCAGCGGTAAAGGCGAATTTGAACAGCGTCCGGAAGAAGCCGAAAAAACCGC
CGCAGCACCGGGCGAACGCCTGCCTAGAGATCCGATTACCGGTCAGCTGAGTAGCAGTATT
GAACCGGAATTTTTTGGTGATACCATTCTGGTGAATGGTAAAGCCTGGCCGGTTCTGGAAG
TGGAACCGCGCAAATATCGCCTGCGCGTTCTGAATGGTAGCAATAGCCGTTTTTATCGTCTG
ACCCTGAGCAGCGGTCAGGAATTTCAGCAGATTGGTAGCGATGGCGGTTTTCTGAATGCCC
CGGTTGCCCGCCGTGAACTGCTGCTGGCCCCGGCAGAACGCGCCGATGTGATTGTGGATTT
TGCAGATCCGAAACTGGCCGGTCAGACCATTGTGGTTCGCAATGATGCCCGCACCCCGTTT
CCGAATGGCCATGCAGTTGATCCGGCAACCACCGGTCAGGTTATGGCATTTCGCGTTAGCA
AACCGATGAGTGCAACCGCCGATGCAACCCTGCCGGCCAGTCTGCGCGCCCCTTTAGCAA
AACTGCCGGGTCTGCGTGCCCGCGTTCGTCAGCTGCTGCTGGCGGAAATTAAAGATGAATT
TGGCCGCATTAAAACCATGCTGGGCACCGTTGAACATGGCGCACTGGGCTGGGATGCCCCG
ATTAGCGAAACCCCGCGCCGTAATGATGTGGAAATTTGGAGCGTTGTGAATGCCACCCCGG
ATGCCCATCCGATGCATCTGCATATGGTTTTTTTTCAGGTTCTGGATCGCCAGAAATATGATG
CCGAAAAATTTGAAGCAGGCAAACCGGCCACCCTGCGCCTGACCGGTACCGCAATGGCAC
CGCCGGCCGGTGAACGCGGTTGGAAAGATACCGTTATTATGCGTCCGGGTGAAGTGACCCG
TGTTATTGCCCGCTTTGATCTGCCGGGTTTATATGTGTGGCATTGCCATATTCTGGAACATGA

AGATCATGAAATGATGCGCCCGTATCGTGTTCTGCCGTAAaagctt
MDNVLVPDTTTYPGTDYYHVRMVQVQQDLGLIDPETRKPLRTTVWAYGAANRTATYPG PTIVAHSTLATNRQPGKPVKVRWENALPDKHLLPVDTTVHCGPDARQQHSHCRPFVRTVTHL HGGHVPDHSDGYPEAWFSPGFREKGPLWSREVYDYPNDQEAATLWYHDHAMGITRLNVYAG LAGFYLVRDDNEQRLQRDGRLPAPAYDVPLVIQDRSFRRDGSLFYTSGKGEFEQRPEEAEKTA AAPGERLPRDPITGQLSSSIEPEFFGDTILVNGKAWPVLEVEPRKYRLRVLNGSNSRFYRLTLSS GQEFQQIGSDGGFLNAPVARRELLLAPAERADVIVDFADPKLAGQTIVVRNDARTPFPNGHAV DPATTGQVMAFRVSKPMSATADATLPASLRAPLAKLPGLRARVRQLLLAEIKDEFGRIKTMLGT VEHGALGWDAPISETPRRNDVEIWSVVNATPDAHPMHLHMVFFQVLDRQKYDAEKFEAGKP ATLRLTGTAMAPPAGERGWKDTVIMRPGEVTRVIARFDLPGLYVWHCHILEHEDHEMMRPYR VLP
Bcece
ggatccATGGGAAAATGTGTAAAGCCATCAGATCCCAACTCTATCCCTAAATTCATGGAcC CGCTACCGAAGCCGAGCGTTGCAGTCCCGGTTCAGCACGATGACTACCCGGATGGCAGCTA TTACGAGCTGGAGATGAAAGAAGGTTTGCACCGCTACCATCAAAACTTTCCGGAAACCAA GATTTGGGGTTACGATGGTCTGATACCTGGTCCGACCATCGAAGCGCGTAAAGACAAAACC ACCTATGTTAAATATTTAAATAATCTGCCGGATGAGCACTTCCTGCCTCTGGATCGCACCTTG CACAGCTCCATCGATACCGCAGATGTGCGCACCGTCGTGCACCTGCACGGCGCTAAAGTTG ATTGGGAAAGCGACGGTCATCCGGAAGCATGGTACACCAAAAACTACGAGATGACCGGTC CGACCTTTCGTCGTCAGGTTCACGCGTATACCAATCACCAGCCGGGTGCCACCCTGTGGTAT CATGACCACGCTATGAGCTTGACCCGTCTTAACGTGTATTCTGGCCTGGCAGGCTTTTACCT GCTGCGTGATGCCCTGGAAGATCGTTTGAGACTGCCGAGTGGTAAGTACGAGATTCCGATG ATGATTCAAGATCGCAGCTTTAATGATGATGGCAGCTTGTTCTACCCGGACACCCCGCCGTT CCCGGTTACGGTCAATCCGAGCATCACCCCCGGCGTTTTAGGTGATACGATTGCGGTCAAC GGCAAGCTGTGGCCGTATCTGAACGTGGAGCCGCGTAAGTATCGTTTTCGTGTTCTCAACG CGAGCAATCGTCGTGGTTACTCCCTGTCTCTGAGCAACGGTGGTACGATCCATCAGATTGGC ACGGACGGCGGTTTGTTGGAGGCTCCGGCGGAGTTGACCTCGTTCGATCTGCTGCCGGCTG AGCGCACAGACATTATTATCGACTTTTCTACGATGCAAGGTCAGACCATCACCTTGTTGAAC AACGATCCAGAATTCCCGCTGAATGAGCACACCTCCGTGGTTATGCAGTTCAGCGTTTGTCT GCCGTTAGATGGTGAAGATGACAGCCAGATTCCGGAGACCTTATATCCGGAAATGGCGCTG CACACCGAACACGCGCATATCGTGCGTAACCTGCCGTTGACCGCGACCACCGACGAATATG GTCGTCCGATGCTGATGTTGAACGACCATATGTATCATGACCCGGCTAGCGAACGCCCGTCG CTCGACAGCATCGAGATCTGGAATTTCATCAACACCACCCCGATCCAACACCCGATTCACCT GCACCTGATTCAATTTAAAATCCTGGAGCGCCGTCCGTTTGATGTTGATGTGTTCACCGAAA CCGGTGAAATTGTTTTTACCGGTCCAGCCGAGCCGCCACGTGAATACGAGCGTGGATGGAA GGATGTTGTGCGCGCTGATATCGGGATGGTTACCTCCATTGCTATGCATTGGAAGGACTTCA CCGGAAACTATATTTGGCATTGTCATTTCCTGGAACACGAAGATCACGACATGATGCGTCCG ATTACCATCATCGAAAATACGCATCCAGTTCAGCTGCCGCATGCTGACGAGGCCCCGGCGG AGGTTCCAGCGCCAGCGGAACCGCCAACCGACACCGTGACACCGGAGGACCCACAGGTG ACCGACGAAACAGCGGAGGAAACGGTTAACGAAGTTATTGAAGGTACGGATGATACCACA GCGACCACCGACACGGAAGGTAATGGTGATAGCGACGCCGGCACGGTGGCAGAGGGGACT GACGATTCTAACGCAGGGACTCCCGCGGAGGGCGCCGATGATTCCGACGCGGGGACCGCG GTGGAAGGAGCCGACGACTCCGGTGCGGGCACAACGGTGGAAGGCGCCGATGACTCCGG GGCGGCAACGACCGCGGAGGGCACAGACGACTCCGACACCGGTACCACGGTTGAGGGTG

CAGATGACAGCGGCGCAGGCACGCCGGCGGAAGGTGCGGTGGATTCAGATGCGGGCACG
ACCGTGGAGGGTGCCGACGACTCTGGCGCCGGCACCCCGGCCGAGGGTGCGGTCGATTCC
GACGCCGGTACCACTGTGGAGGGCACCAATGACACTGACGCTGGCACCACGGCGGAGGGC
ACTGACGACTCTGGCGCTGGTACCACCGTGGAAGGTACGGACGACTCCGATGCAGGCACT
ACCGCAGAAGGTACCGACGATTCCGGCGCGGGTACCACCGTCGAGGGCACGGACGTGAGC
GACGCGGGTACGGCTGTGGGTAGCACCGATTGGAAGAGCACCATGGCGATCCAGGGCTAC
GAAGCGATTGAAACGCAAAACGCGACTGACCGCGCAGAGACCGGCATGGCACCCGCTGA
GGGCACTGCGCCCGACGGCACCGACGACACCGGTGCAGGCACCACCGTAGAGGGTACCG
ATGTAAGCGACGCTGGTACTGCGGTGGGTTCAACCGACTGGAAGTCGACCATGGCCATCCA
AGGTTACGAGGCAATCGAGACCCAAAACAGCCCGGGTAGAGCAGGTACTGGTATGGCTCC
GACCTACAAACAGAAACCGTGCAACTGCCCGAAGTGCCGTTACGCACGGAAGAAAAAGA
AAGGTAAATAAaagctt

MGKCVKPSDPNSIPKFMDPLPKPSVAVPVQHDDYPDGSYYELEMKEGLHRYHQNFPETKI
WGYDGLIPGPTIEARKDKTTYVKYLNNLPDEHFLPLDRTLHSSIDTADVRTVVHLHGAKVDW
ESDGHPEAWYTKNYEMTGPTFRRQVHAYTNHQPGATLWYHDHAMSLTRLNVYSGLAGFYLL
RDALEDRLRLPSGKYEIPMMIQDRSFNDDGSLFYPDTPPFPVTVNPSITPGVLGDTIAVNGKLW
PYLNVEPRKYRFRVLNASNRRGYSLSLSNGGTIHQIGTDGGLLEAPAELTSFDLLPAERTDIIIDF
STMQGQTITLLNNDPEFPLNEHTSVVMQFSVCLPLDGEDDSQIPETLYPEMALHTEHAHIVRNL
PLTATTDEYGRPMLMLNDHMYHDPASERPSLDSIEIWNFINTTPIQHPIHLHLIQFKILERRPFDV
DVFTETGEIVFTGPAEPPREYERGWKDVVRADIGMVTSIAMHWKDFTGNYIWHCHFLEHEDH
DMMRPITIIENTHPVQLPHADEAPAEVPAPAEPPTDTVTPEDPQVTDETAEETVNEVIEGTDDTT
ATTDTEGNGDSDAGTVAEGTDDSNAGTPAEGADDSDAGTAVEGADDSGAGTTVEGADDSGA
ATTAEGTDDSDTGTTVEGADDSGAGTPAEGAVDSDAGTTVEGADDSGAGTPAEGAVDSDAGT
TVEGTNDTDAGTTAEGTDDSGAGTTVEGTDDSDAGTTAEGTDDSGAGTTVEGTDVSDAGTAV
GSTDWKSTMAIQGYEAIETQNATDRAETGMAPAEGTAPDGTDDTGAGTTVEGTDVSDAGTAV
GSTDWKSTMAIQGYEAIETQNSPGRAGTGMAPTYKQKPCNCPKCRYARKKKKGK

Scla

ggatccGCTCCACCCGGAGGGCTACCTGCAGTAAGGGACCCTCGTTACCTGGCTCCGATT
GCGGATGCGAGAGCGGTGCCCCGTTTCGTGACGCCGCTGCCGCGCCCAAGCCGTATTGATC
TCTCGGATGGCGGCTCGCGCACCTTGGTGATGGCGCCGGTTACCCAAGACGTGGTGGGTGC
AGGCCTCGGTTTGCGCACTCCGGTGTGGGGTTTCGGCGCGGTTGGTGGTCCGGCGGGCGC
TAGGCCGAGCTGGCCTGGCCCGACGCTGGTTGCGCGTTCCGGCCGTCCGGTGCGTGTTCGT
TGGCGTAATGCGCTGCCGTTCCGTCATCTACTACCGCTGGACACCACTGTTCATTGGGCGTT
CAGCGGAACGGGCCGTAGCATCGAGCACGATGGTGTTCCGACCATTGTGCATCTGCACGGT
GGACACAGCGACGCTGGTTCCGATGGTCACCCGGACGCCTGGTACACCCCGGGTGGTGCG
AGAGGTCCGCGCTTCACCGGCACCCGTCCGACCTATGACAACAGCCAGGAGGCGGCGACG
TTATGGTATCATGATCACACCTTAGGGCTAACTCGTCTGAACGTGTATGCGGGTTTGGCTGG
CTTCTACCTGTTGCGTGACGACCGCGAACTGTCTCTGATCGAACGTGGCGCCCTTCCGTCT
GGTCCGTACGAGGTGGAGCTGGCGATCCAGGATCGAGACCTGCGTCCGGACGGCAGTCTG
GCGTTTCCGGACGCTCCGGCGGCAGCTCCAGATTGGCCGGGTGGCCCGTCTGTACTGCCGG
AGTTCTTTGGCAGCGTTATTGTTGTCAACGGTGCTGCATGGCCGGTGCTGGACGTAGAGCC
GCGTCGCTACCGTTTGCGTCTGCTGAATGCGAGCAATTCACGTTTCTATCGTCTTAGCGCAC
ATGGTGTTCGTCCGGTGCCGGTTACCCAGATTGGTGCTGACGGCGGTTTTCTGGACCGCCC
GGCAGCCCTGGACCGTCCGTTGCTCTTAGCCCCGGGCGAACGTGCGGATCTGATCGTTGAT

TTTCGCGGCCACGAGGGTGCACTTATCGATCTGACGAACGATGCGCCAGTCCCGTTTCCGG
ACGGCGATCCGGTCGCCCCTCCGGCTGACCGTGTGCTGCGCTTCCGCGTAAGCCTGCCCTA
TGACCGTCGTGTTCCGGAAGCAGTTCCTCCGCGCACCCTGAGAGAAACCCCGTTCCGTGCC
GGTGGCGCCCCAGCGCGCACCCGCCGTTTGGTGTTAACCGAGCACATGGATAGATACGACC
GTCCGAAGCCGATGTTGGGCACGGTCGAGCGCGGCATGCTGGAGTGGATGGAACCGGTTA
CTGAAACCCCGGCGCTGCACAGCACCGAGATCTGGGAATTCTACAACACCACAGATGACG
CCCACCCGGTCCATCTCCACTTGGTCCAATTTCAGATCCTGGACCGTGCGCCGTTTACCGCA
GTTCAAGATCCGGTGACCGGTGCGTTGTCCGACATCGCCACCGGATCGCGGCGCCCGCCAG
GCGCTGATGAAAACGGTCCGAAAGATACCGTTAGAGCGCTGCCTGGTGAAGTTACCCGTCT
GCGCGCGGTGTTTGATAAACCGGGTGTATTTGTGTGGCATTGCCATATTCTGGACCACGAAG
ATCACGAGATGATGCGTGAATATGAGGTGGTCGACGACGGTTAAaagctt

APPGGLPAVRDPRYLAPIADARAVPRFVTPLPRPSRIDLSDGGSRTLVMAPVTQDVVGAGL
GLRTPVWGFGAVGGPAGARPSWPGPTLVARSGRPVRVRWRNALPFRHLLPLDTTVHWAFSGT
GRSIEHDGVPTIVHLHGGHSDAGSDGHPDAWYTPGGARGPRFTGTRPTYDNSQEAATLWYHD
HTLGLTRLNVYAGLAGFYLLRDDRELSLIERGALPSGPYEVELAIQDRDLRPDGSLAFPDAPAA
APDWPGGPSVLPEFFGSVIVVNGAAWPVLDVEPRRYRLRLLNASNSRFYRLSAHGVRPVPVTQ
IGADGGFLDRPAALDRPLLLAPGERADLIVDFRGHEGALIDLTNDAPVPFPDGDPVAPPADRVL
RFRVSLPYDRRVPEAVPPRTLRETPFRAGGAPARTRRLVLTEHMDRYDRPKPMLGTVERGMLE
WMEPVTETPALHSTEIWEFYNTTDDAHPVHLHLVQFQILDRAPFTAVQDPVTGALSDIATGSRR
PPGADENGPKDTVRALPGEVTRLRAVFDKPGVFVWHCHILDHEDHEMMREYEVVDDG

Arose

ggatccATGCAGCTGACCCCGTATAAAGATAGTCTGACCGTGCCGAAAAAACTGATTATTA
AACGCGGTAAACTGAAAATTGCCATGGTTCGCACCCGCCTGCGCCTGCATAGTGAACTGCC
GAATACCACCGCCCTGTGGACCTATGAAGGCAGTGCCCCGGGCCCGATTATTGAAGTGCGT
CGCGGCCAGAAACTGACCGTGGAATGGAAAAATCAGCTGACCGGCACCATTCCGCTGACC
GCCGCACGCAGCAAACAGGTTCCGGGCGATGAAGATCTGTGGCATAGCAATGTGGCAGGT
ATTGCCCTGAGCAGTGATGCAGATGAAGTGATGATTACCCCGCATATTGATGATCTGCATGC
ACATACCGTTGTTCATCTGCATGGTGGTAAAACCAGCCCGGATAGTGATGGTTGGACCGATA
ATGTTCATCATCTGGGTCAGACCCAGGCATGCGTTTATGAAAATGATATGCGTAGTACCCTG
CTGTGGTATCATGATCATGCCATGGGCGTTACCCGCTTTAATGTTTTTAGTGGTCTGGCCGGC
TTTTATATTATTCGTGATGCAGATGATGACAAAGTTATGAGTGCACTGCGTGCAAGTCATGGT
CTGACCAAAAAAGCACGCAGCCCGAAAACCCCGCTGGAACTGCCGCTGGTTATTCAGGAT
CGTAATTTTCTGACCAAAGCAAATGGCGATCTGACCGGCGAACTGCTGCATATTGTTCAGG
ATCTGCCGCTGAATGCAGCCGGCGATAAAGGTCCGATGGAATTTTTTGGCCCGTATACCCTG
GTGAATGGTACCGTGTGGCCGCATGTTGATGTGCGCCCGGGCCCGTATCGTCTGCGCTTACT
GAATGGCAGCAATGCCCGTACCTATGCACTGGCCCTGGTGGATCAGAATAATAGCCCGGTTC
CGAATGTTACCATGCAGCAGATTGGTACCGATGGCGGCCTGCTGGGCGCCCCTAGACCTGT
TGAACGTATTGTGATTGCACCGGCAGAACGCGTTGATCTGATTATTGATTTTACCAGCGTGG
CCGCAGGTACCAAACTGCGCTTTCTGAATACCGCCCTGAGCCCGTTTGGCAATCAGTATGT
GCCGAAACCGGGTGATACCGCACCGAGCGATCTGCTGGTGCTGCCGGAAGTTATGGAATTT
CGTGTGCAGGGTAAAAAATGCCAGCGCCTGGATCTGCCGGCCCCGTTAAGCAAATTTAAAG
CCCTGACCCATGGTGATCTGCCGCATCATCATAATCATCGCCTGGTTGCACTGGTGGAAGAT
ACCAATACCTTTGGTATGCTGACCTTTCTGGAACTGGAAAAAGTGGGCCCGGCCCAGACCC
CGACCCCTGACGGTATGCTGAATCTGCCGCTGGATATTCATCCGGATGATCCGGCCAATCCG

CTGCCGAGTGTGGATTATCGTATTACCGCCCGCATGTTTCGTGATGGTAATCCGTGGATGGT
GGCCGAAAATGCAACCGAAGTGTGGAATATTATTAATCTGACCGGTGATACCCATCCGTTTC
ATATTCATCTGGTGCAGTTTCAGATGCTGCGTCGCACCGGTTTTGATCCGCAGAATTTTGAA
TATAATGACGGTAAAATCGCGAATCAGGGTCCGATTATTAGTAAAGGCACCAGTCAGCCGCT
GGAAATTGGTTGGAAAGATACCGTGCGCGTGGACCCTGGTGAACTGGTTAGCATTGCAGTT
CCGTTTGAAGGCTTTAGCGGCCAGTATATGTATCATTGTCATATTCTGGAACACGAAGATCAT
GAAATGATGCGTCCGTTTGTTGTGATGAGCGAAGAAACCATGCCGTTTATGCCGAGTCATG
GCCATAGTAGCGGCCATGATCATAATCATTAAaagctt

MQLTPYKDSLTVPKKLIIKRGKLKIAMVRTRLRLHSELPNTTALWTYEGSAPGPIIEVRRG
QKLTVEWKNQLTGTIPLTAARSKQVPGDEDLWHSNVAGIALSSDADEVMITPHIDDLHAHTVV
HLHGGKTSPDSDGWTDNVHHLGQTQACVYENDMRSTLLWYHDHAMGVTRFNVFSGLAGFY
IIRDADDDKVMSALRASHGLTKKARSPKTPLELPLVIQDRNFLTKANGDLTGELLHIVQDLPLN
AAGDKGPMEFFGPYTLVNGTVWPHVDVRPGPYRLRLLNGSNARTYALALVDQNNSPVPNVT
MQQIGTDGGLLGAPRPVERIVIAPAERVDLIIDFTSVAAGTKLRFLNTALSPFGNQYVPKPGDTA
PSDLLVLPEVMEFRVQGKKCQRLDLPAPLSKFKALTHGDLPHHHNHRLVALVEDTNTFGMLTF
LELEKVGPAQTPTPDGMLNLPLDIHPDDPANPLPSVDYRITARMFRDGNPWMVAENATEVWNI
INLTGDTHPFHIHLVQFQMLRRTGFDPQNFEYNDGKIANQGPIISKGTSQPLEIGWKDTVRVDPG
ELVSIAVPFEGFSGQYMYHCHILEHEDHEMMRPFVVMSEETMPFMPSHGHSSGHDHNH

Cmeth

ggatccGCTGCAGGTACCCCGAATATTACCAAATTTGTGGACCCTCTGCTGAGTAGTATTC
CGAAACTGATTCCGGTGGCCAATGATGCATATCCGGGTGCCGATTATTATGAAATTGAAATG
AAACCGGGCCGCTTTCAGTTTAGTAGCCAGCTGCCGGTTACCGCCCTGGCAGGCATTAATA
CCACCTGGGGTTATGGTGGTCGTGATCTGGCAAATAAAGGTCCGCTGGTGTTTCTGGGTTAT
CTGGGTCCGACCATTGAAGCACAGAAAGGTAAAACCGTTGTGGTGAAATTTATTAATAACC
TGCCGACCACCCATCCGCTGCAGGCCAGCATTGATCCGACCGTGCCGGACCCTGCAATGTA
TAGCGATCTGTATGATACCAGCACCAATCCGCCGACCCCGATTCATATTGGCCGCACCACCC
CGCATCTGCATGGCGGTTTTACCGCCCCGCAGTTTGATGGCCATCCGCATAGTTGGTTTACC
CCGATTAGTGATGGCGGTAAAGGTAGTCATTATGCCACCCTGAGCGGTGCAGCAGCCAATG
AAGCCATTTTTGTTTTTAGCAATCAGCAGCCGGCAACCAATCTGTGGTATCATGATCATGCC
ATGGGCATTACCCGCCTGAATGTGTATGCCGGTCTGGCCGGTCTGTATTTTGTTCGCGATCA
GTATGATACCGGCACCAGCAATCCGCCGACACCGGGTCTGAATCTGCCGGCAGGTAATTAT
GAAGTTCCGCTGGTTCTGCAGGATAAACAGTTTAATGCAGATGGTACCCTGTTTTATCCGAC
CGTTGGCATTACCGCAGTTCATCCGATTTGGATGCCGGAATTTTTTGGTGATACCCCGGTGG
TGAATGGTACCGCCTATCCGTTTCTGAGTGTTGAACCGCGCCGTTATCGTTTTCGTATTCTGA
ATGGTAGTCAGGCCCGCTTTTATAATCTGTGGTTTGATAATGGTCTGGCACCGATTCCGTTTC
ATCTGATTGGCATGGAACAGAGCCTGGTGCCGACCCCGGTTGCAATGACCAAACTGCTGAT
TGCACCGGGTGAACGCGCCGATATTATTGTGGATTTTGCCGGTATGCAGAATCAGATGCTGA
CCCTGAAAAATAATGCCAAAGCCCCGTATCCGGGCGGCAAAGGTGGCTTTGGTCAGATTAT
GCAGGTGCGTGTGAATCTGCCGATGACCGGTACCGATACCACCACCCCGGCAGCCAATCTG
GTGCTGCCGCAGGTTCCGCGTCTGACCGGCGCTACCATTATGCGTGAAATTGTGATGAAAG
AAGATGTTGATCCGGTTACCGGTCTGCCGATTGATATGAAACTGAATGGCAAATGGTTTAAT
GATCTGCCGATTGACGAAACCCCGACCCAGAATGCCACCGAAGTTTGGCAGTTTATTAATCT
GACCGTGGATGCACATCCGATGCATATGCATCTGGTTAAATTTCAGGTTAGCGATCGCCAGC
CGTTTGATGCCAAAGCCTATACCACCGCATGGCTGAAATATCAGGCAGGTCTGGGCAGCAA

ACCGATTCTGGCAAATTTTCTGACCAAAGGCCCGGCCGTGCTGCCGCCTCCTGAAGAAATG
GGCTGGAAAGATACCGCCAAAAGCTATCCGGGCGAAATTCTGCGCGTTATTGCAAAATTTG
AACTGCCGGATGCAGATAGCAATATTCCGGGTAGCGGTACCCAGCTGCCGGCAGAATATGTT
TATCATTGTCATATTCTGGAGCATGAAGAAAATGAAATGATGCGCCCGTTTACCGTGGGTTA
Aaagctt

GSAAGTPNITKFVDPLLSSIPKLIPVANDAYPGADYYEIEMKPGRFQFSSQLPVTALAGINT
TWGYGGRDLANKGPLVFLGYLGPTIEAQKGKTVVVKFINNLPTTHPLQASIDPTVPDPAMYSD
LYDTSTNPPTPIHIGRTTPHLHGGFTAPQFDGHPHSWFTPISDGGKGSHYATLSGAAANEAIFVF
SNQQPATNLWYHDHAMGITRLNVYAGLAGLYFVRDQYDTGTSNPPTPGLNLPAGNYEVPLVL
QDKQFNADGTLFYPTVGITAVHPIWMPEFFGDTPVVNGTAYPFLSVEPRRYRFRILNGSQARFY
NLWFDNGLAPIPFHLIGMEQSLVPTPVAMTKLLIAPGERADIIVDFAGMQNQMLTLKNNAKAP
YPGGKGGFGQIMQVRVNLPMTGTDTTTPAANLVLPQVPRLTGATIMREIVMKEDVDPVTGLPI
DMKLNGKWFNDLPIDETPTQNATEVWQFINLTVDAHPMHMHLVKFQVSDRQPFDAKAYTTA
WLKYQAGLGSKPILANFLTKGPAVLPPPEEMGWKDTAKSYPGEILRVIAKFELPDADSNIPGSG
TQLPAEYVYHCHILEHEENEMMRPFTVG

Cohnella

ggatccATGACCGTGAATCCGGCCGATCCGGAAACCATTCCGAAATTTGTTGATCCGCTGC
CGATTCCGCCGATTCTGAAACCGGTTCGCAAAATTAATGGCAAACCGTATTATGAAGTGCGC
ATGCGTCAGACCCAGCAGCAGGTTCATCGCGATTTTCCGGTTACCACCCTGTGGGGTTATAA
TGGCATGTGGCCGGGTCCGACCATTGCCGCACGTCGTAATTGCCCGATTAAAGTGCATTGG
ATTAATGAACTGCCGACCACCAGTCTGCTGACCGTGGATCATAGCATTTTTGGTGCAGAAGC
ACTGTTTAGTGGCAATGGTACCCAGAATTTTCGCCCGGATGTTCGTAATGTGGTTCATCTGC
ATGGTATTAATGCCCCGGCAGAATATGATGGTACCCCGGATCAGTGGTATACCCCGGGTCTG
ACCCAGATTGGTCCGCTGTTTAAAACCGATGTGTTTGAATATCCGAATCGCCAGCCGGCAGC
ACAGCTGCATTATCATGATCATGCAATTGGCATTACCCGTCTGAATGTGTATAGCGGCCTGCA
GGGCTTTTATTTTATTCGCGATGCAGTTGAAGAAAAACTGCCGCTGCCGAAAATTCCGTATG
AAATTCCGATTCTGATTACCGATCGTCAGTTTAATCCGGATGGTAGTCTGTTTTATCCGGAAC
ATTGGGAACCGCAGTTTCTGGCAAATACCATTGCAGTTAATGGCAAACTGTGGCCGTATCTG
GATGTGGAACCGCGTCGTTATCGCTTTCGCATTCTGAATGGTAGCAATAGCCGTTTTTATCA
GCTGCGTCTGGACCCTGTGCTGACCTTTCATGTTATTGGTACCGATGATGGTCTGATGGAAC
GTCCGGCAGCAGTGGAAGAATTTCTGCTGGCCCCGGGCGAACGTGTGGATGTTATTGTTGA
TTTTACCGGTCAGCGCGGCAAAACCTTTACCATGACCAATAGCGCACTGGCCCCGTTTGAT
GTTGGCCTGCCGCCGAATCCGAATACCGATGGCCTGATTATGCAGTTTCGCGTGAATCAGCC
GATGAGCGGCAAAGATCATAGTCGTATTCCGAAACGCCTGGCACAGCTGAAACGCCTGAAT
CCGAATAAAGTGGCAAAAATTCGTGATGTTACCCTGAATGATGGCACCGAACGTGAAGGTG
AAGAAGAACGCGCAATTGTGCTGCTGGGTACCCGTAATCTGACCGGTCGTACCCAGCCGTA
TCTGTGGGGTGATAGCACCACCGAAAAACCGGTTCTGGGTAGCACCGAAATTTGGCGTCTG
ATTAATACCAGCCCGGGCACCCATCCGATTCATCTGCATCTGGTGCGCTTTCTGATTCTGGAT
CGTCAGCCGTATGATGTGGAACATTTTGAAAAAACCGGTGAACTGGTTTTTACCGGCCCGC
GTGTGCCGCCGGCAACCTATGAACGTGGTTATAAAGATGTGGTGCGCGCAAATCCGGGTGA
AGTTACCCGCATTATTGCACGTTTTGTTGATTATGCCGGTCTGTATGTGTGGCATTGTCATATT
CTGGAACATGAAGAAAATGAAATGATGCGTCGTATGGAAGTGGTGCGTCCGGGCTGTCGTA
CCAGCGAACCGGCAGCAAGCCGCAGTTGGGCCGCAGTTCCGCGTCGCCAGAATCGCCATG
TTAATCGTCGCCGCGGTGTGCATGAAATTCGCAGCAGCATTAAACGCCGTTGCCGTCGTCG

CTAAaagctt
MTVNPADPETIPKFVDPLPIPPILKPVRKINGKPYYEVRMRQTQQQVHRDFPVTTLWGYN GMWPGPTIAARRNCPIKVHWINELPTTSLLTVDHSIFGAEALFSGNGTQNFRPDVRNVVHLHGI NAPAEYDGTPDQWYTPGLTQIGPLFKTDVFEYPNRQPAAQLHYHDHAIGITRLNVYSGLQGFY FIRDAVEEKLPLPKIPYEIPILITDRQFNPDGSLFYPEHWEPQFLANTIAVNGKLWPYLDVEPRRY RFRILNGSNSRFYQLRLDPVLTFHVIGTDDGLMERPAAVEEFLLAPGERVDVIVDFTGQRGKTF TMTNSALAPFDVGLPPNPNTDGLIMQFRVNQPMSGKDHSRIPKRLAQLKRLNPNKVAKIRDVT LNDGTEREGEEERAIVLLGTRNLTGRTQPYLWGDSTTEKPVLGSTEIWRLINTSPGTHPIHLHLV RFLILDRQPYDVEHFEKTGELVFTGPRVPPATYERGYKDVVRANPGEVTRIIARFVDYAGLYVW HCHILEHEENEMMRRMEVVRPGCRTSEPAASRSWAAVPRRQNRHVNRRRGVHEIRSSIKRRCR RR

*All sequences were codon optimized and the signal peptides were removed.

Table S4 The evaluation for different positive threshold

Positive threshold	Accuracy	Precision	Recall	F1-score
50%	1.0000	1.0000	1.0000	1.0000
90%	1.0000	1.0000	1.0000	1.0000
99%	0.9990	1.0000	0.9667	0.9831

Table S5 The five-fold evaluation results of the MCOs model

Fold	Accuracy	Precision	Recall	F1-score	ROC-AUC
1	1.0000	1.0000	1.0000	1.0000	1.0000
2	0.9981	1.0000	0.9310	0.9643	0.9861
3	0.9981	1.0000	0.9310	0.9643	0.9492
4	0.9990	1.0000	0.9655	0.9825	0.9977
5	0.9990	1.0000	0.9655	0.9825	0.9858
Avg.	0.9988	1.0000	0.9584	0.9787	0.9838

Table S6 Characterization of the multicopper oxidases

enzyme	Specific activity (U/mg)	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1}/mM)	Halt-life $t_{1/2}$ 80°C (min)
Eclac (query)	0.41±0.10	6.91±1.32	1.50±0.28	0.22	23.1±1.2
Sulfur	13.51±0.74	0.94±0.11	17.48±1.88	18.60	156.9±9.0
Salvi	0.61±0.13	6.08±0.77	2.03±0.13	0.34	14.6±1.5
Ypes	0.47±0.06	12.16±3.74	4.76±0.97	0.37	9.6±0.8
Psoli	0.20±0.01	9.78±2.03	0.78±0.11	0.08	27.2±1.4
Faest	0.04±0.01	32.53±5.85	0.76±0.12	0.02	16.0±1.1
Mint	0.04±0.01	12.13±2.51	0.69±0.10	0.06	6.0±0.3
Ppht	0.70±0.10	0.54±0.18	0.86±0.20	1.66	44.2±7.0
DSM13(query)	3.12±0.37	0.58±0.05	2.39±0.12	4.12	23.2±1.7
Arose	6.67±0.78	8.81±1.31	13.27±1.76	1.51	10.6±1.2
Scla	13.38±1.23	6.76±1.65	7.30±1.19	1.08	62.2±5.2
Bcece	0.55±0.04	0.33±0.02	1.10±0.02	3.33	68.8±8.3
Cmeth	0.05±0.01	ND	ND	ND	ND
Cohnella	ND	ND	ND	ND	ND
HR03 (query)	1.96±0.09	12.45±1.96	6.22±1.07	0.50	66.3±2.6
Aory	0.37±0.04	1.29±0.35	2.16±0.26	1.95	43.3±4.1
Bfre	2.07±0.04	0.10±0.03	4.46±0.40	47.63	61.1±6.4
Tocean	1.61±0.04	0.71±0.16	2.01±0.46	2.88	65.6±5.1
BMS3	0.001	ND	ND	ND	ND
Talbi	0.006	ND	ND	ND	ND
Acid	ND	ND	ND	ND	ND

ND means not detected.

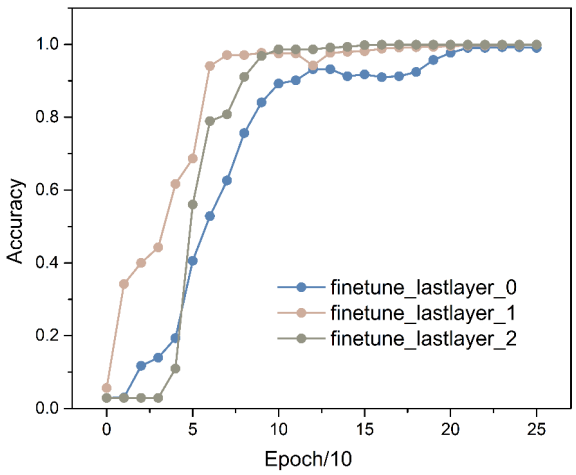


Figure S1 Changes in accuracy during the model fine-tuning process.

After fine-tuning the classification layer of the model, the classification accuracy can approach 1.0. lastlayer_0: last layer; lastlayer_1: last two layers; lastlayer_2: last three layers. Source data are provided as a Source Data file.

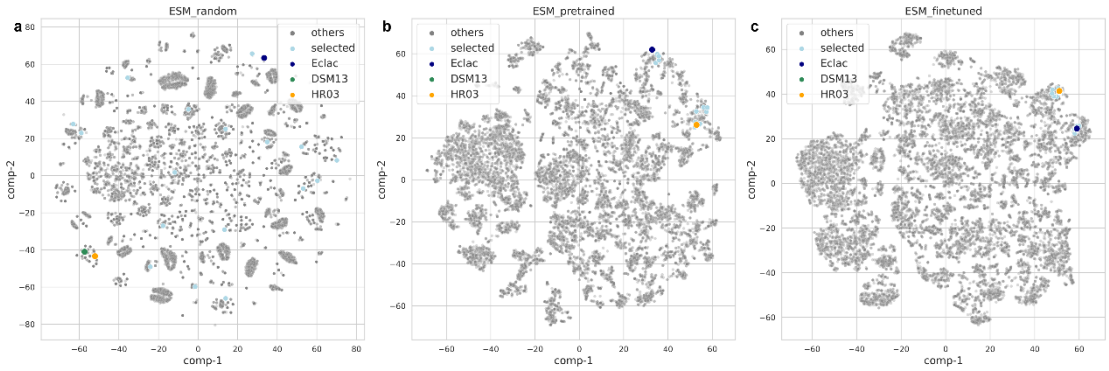


Figure S2 The embeddings of the multicopper oxidases in the ESM-1b model.

a In the random ESM-1b model, all parameters were randomly initialized, resulting in dots being randomly distributed in space. **b** In the pretrained ESM-1b model, the dots of the queried sequences and selected sequences began to cluster. **c** After the model was fine-tuned, the selected sequences became neighbors to the queries. The high-dimensional embeddings were reduced using the t-SNE algorithm. Source data are provided as a Source Data file.

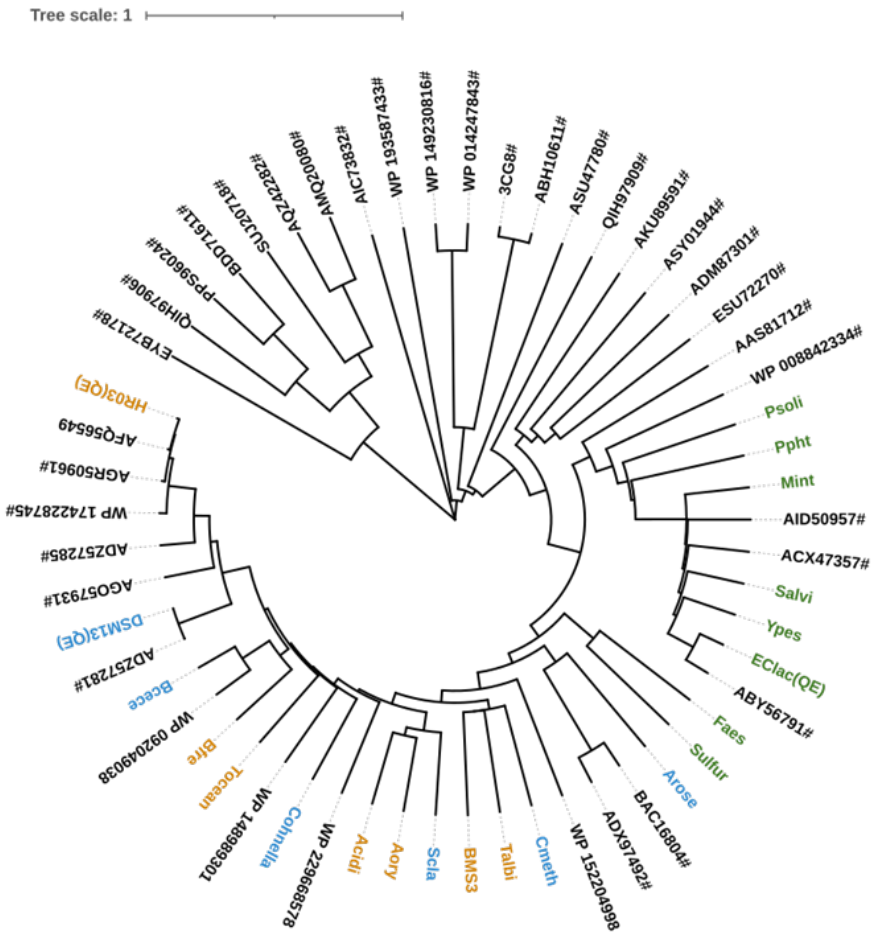


Figure S3 The phylogenetic tree of multicopper oxidases.

Multicopper oxidases retrieved and the enzymes reported (labeled '#') were shown with a UPGMA phylogenetic tree. The multicopper oxidases candidates (18 total, labeling color) are notably diverse compared to the QEs. The protein sequences used for the analysis are provided in Supplementary Table 3. Accession numbers for additional proteins included in the tree are indicated directly on the branches.

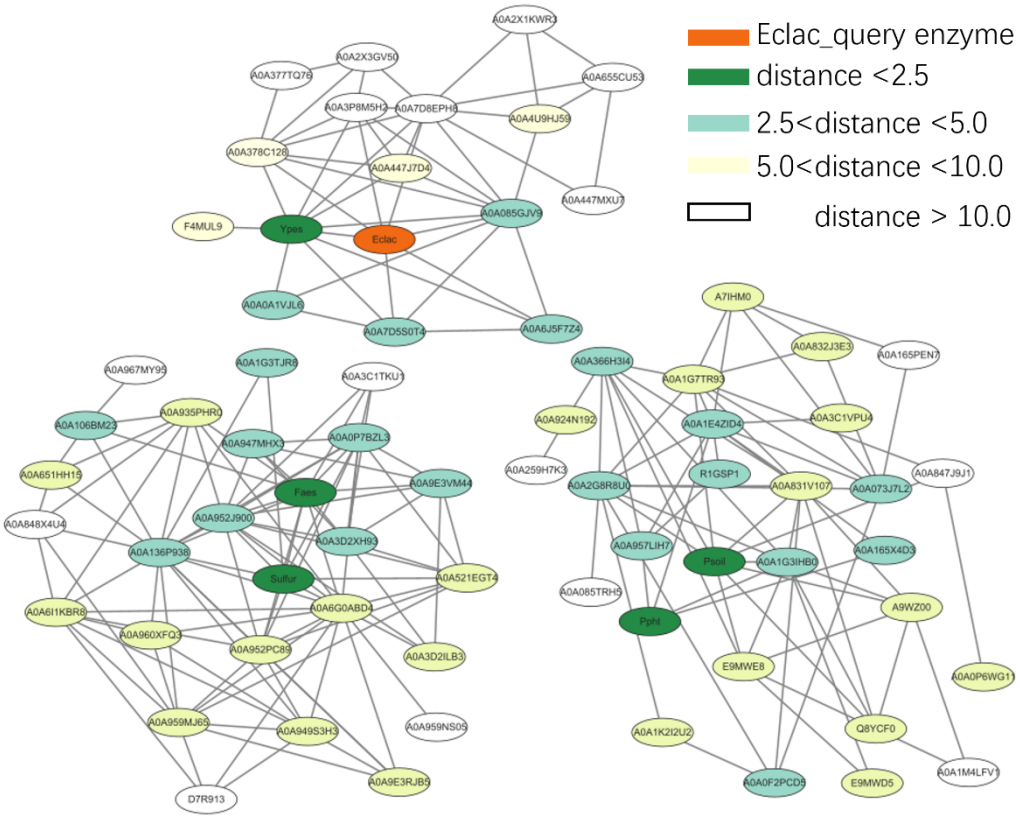


Figure S4 The sequence similarity network (SSN) of Eclac (query enzyme, QE).

The SSN visualizes the relationships between various sequences, emphasizing their Euclidean distances. The network reveals multiple distinct clusters, each representing groups of sequences with closer similarities. Notably, significant variations in Euclidean distances are observed within the same cluster, indicating a spectrum of similarity levels among the sequences, as measured by the Euclidean distance metric. Source data are provided as a Source Data file.

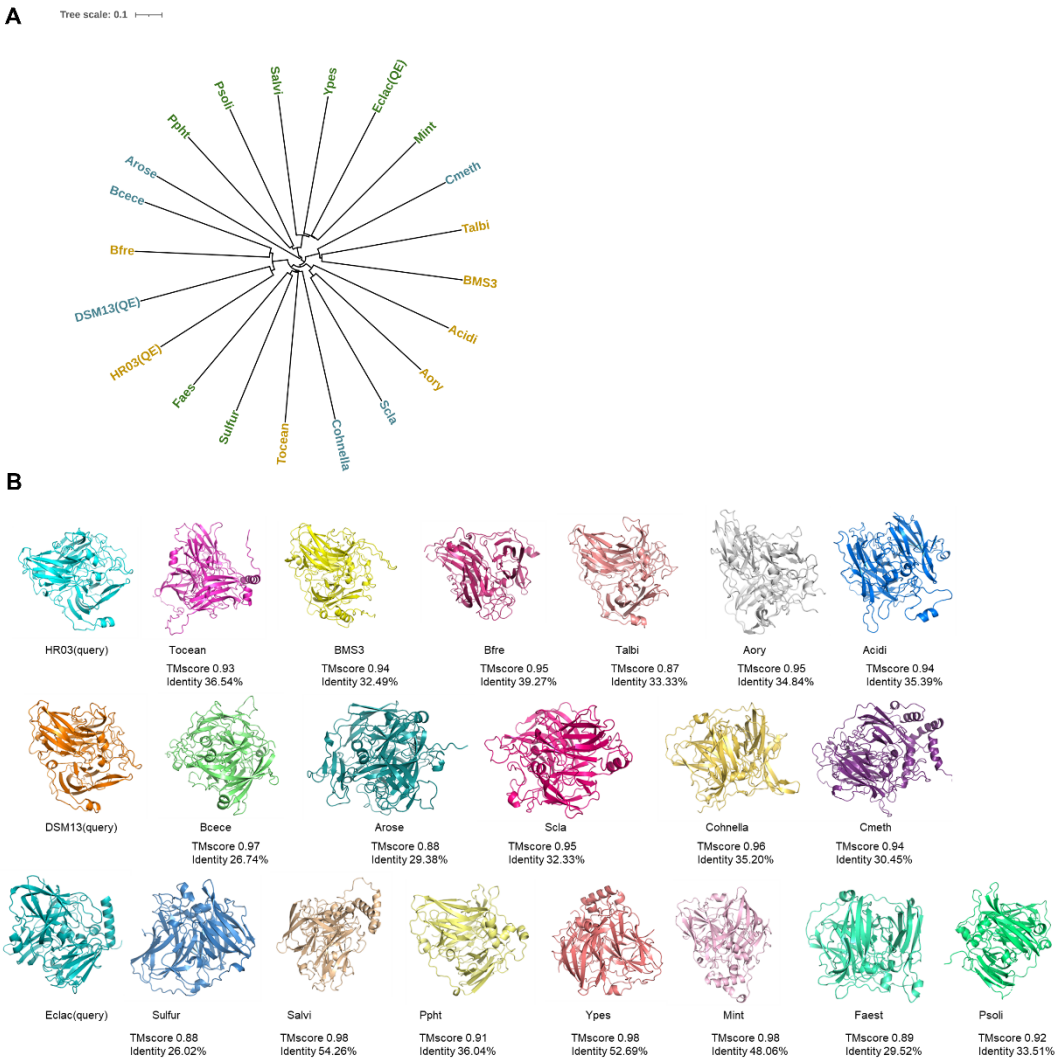


Figure S5 The cluster analysis of all multicopper oxidases characterized in this research.

a Cluster analysis of the protein structure. Each series labeled using different color modes. **b** The AlphaFold2 predicted structures and crystal structures are listed.

Table S7 The enzymes with remote distances tested in this study

Enzyme	Organism	Activity
Eclac(query)	*Escherichia coli* K-12	
A0A447MXU7	*Salmonella enterica I*	ND
A0A377TQ76	*Klebsiella pneumoniae*	ND
A0A2X3GV50	*Klebsiella pneumoniae*	ND
A0A2X1KWR3	*Escherichia coli*	ND
A0A1M4LFV1	*Brucella sp. 10RB9215*	ND

Table S8 The sequences of enzymes with remote distances tested in this Study

A0A447MXU7
ggatccGCTGAACGTCCGGCGCTGCCGATCCCGGATCTGCTGACCGCGGATGCGAGCAACCGT ATGCAGCTGATCGTTAAAGCGGGTCAGAGCACCTTCGCGGGCAAAAACGCGACCACCTGG GGCTACAACGGTAACCTGCTGGGTCCGGCGGTTCAGCTGCACAAAGGTAAAAGCGTGACC GTTGATATCCACAACCAGCTGGCGGAGGATACGACCCTGCACTGGCACGGCCTGGAAATCC CAGGGATCGTGGACGGTGGCCCGCAGGGCATCATTCCGGCGGGTGGCACCCGTACCGTTAC CTTCACCCCGCAGCAGCGTGCGGCAACCTGCTGGATTCACCCGCACAAGCACGGCAAAAC CGGCCGTCAGGTTGCGATGGGTCTGGCGGGTCTGGTTCTGATCGAAGATGATGAAATCCGT AAACTTCGTCCGGCGGAAACCGTTGGCCACCGTCGTTGCACCGGAGATCACAGCGGTCAG ACCCTGCTGCGTCGTTGGCCGGATTAAgaattc
AERPALPIPDLLTADASNRMQLIVKAGQSTFAGKNATTWGYNGNLLGPAVQLHKGKSVTVDIH NQLAEDTTLHWHGLEIPGIVDGGPQGIIPAGGTRTVTFTPQQRAATCWIHPHKHGKTGRQVAM GLAGLVLIEDDEIRKLRPAETVGHRRCTGDHSGQTLLRRWPD*
A0A2X1KWR3
ggatccGCTGAACGCCCGACCCTGCCGATCCCGGACCTGCTGACCACCGATGCTCGTAACCGT ATCCAGCTGACCATCGGCGCGGGCCAGAGCACCTTCGGCGGCAAAACCGCGACCACCTGG GGCTACAACGGCAACCTGCTGGGCCCGGCGGTTAAACTGCAGCGTGGTAAAGCGGTGACC GTTGACATCTACAACCAGCTGACCGAAGAAACCACCCTGCATTGGCACGGCCTGGAAGTT CCGGGTGAAGTTGACGGCGGCCCGCAGGGGATCATCCCGCCGGGCGGTAAACGTAGCGTT ACCCTGAATGTTGATCAGCCGGCGGCTACCTGCTGGTTTCACCCGCACCAGCACGGCAAAA CCGGCCGTCAGGTGGCGATGGGCCTGGCCGGCCTGGTTGTTATCGAGGATGATGAAATCCT GAAGCTGATGCTGCCGAAGCAGTGGGGCATAGATGATGTCCCGGTTATCGTTCAGGATAAA AAATTCAGCGCGGATGGCCAGATCGATTACCAGCTGGACGTTATGACCGCGGCGGTGGGCT GGTTCGCGATCCGTTGTTAAgaattc
AERPTLPIPDLLTTDARNRIQLTIGAGQSTFGGKTATTWGYNGNLLGPAVKLQRGKAVTVDIYN QLTEETTLHWHGLEVPGEVDGGPQGIIPPGGKRSVTLNVDQPAATCWFHPHQHGKTGRQVAM GLAGLVVIEDDEILKLMLPKQWGIDDVPVIVQDKKFSADGQIDYQLDVMTAAVGWFAIRC*
A0A2X3GV50
ggatccATGGGTCTGGCGGGCCTGGTTCTGATTGAAGATGAAGAATCTGGTCGTCTGCTGCTG CCGAAACAGTGGGGCATTGATGATGTTCCGGTTATTGTTCAAGATAAAAAATTCACGGCTGC TGGTGAAATTGATTACCAGCTGGATGTTATGAGCGCAGCGGTTGGTTGGTTCGGTGATACCC TGCTGACCAACGGCGCGCTGTATCCGGAACACGCGGCGCCGAGGGGCTGGCTGCGTCTGC

GTCTGCTGAACGGCTGCAACGCGCGTAGCCTGAACTTCGCGACCAGCGATAAACGTCCGCT GTACGTTGTTGCGTCTGATGGCGGCCTGCTGGCGGAACCGGTTAAAGTTGATGAACTGCCG GTTCTGATGGGCGAACGTTTCGAAGTTCTGGTTGATACCTCCGATGGCAAACCGTTCGATCT GGTTACCCTGCCGGTTAGCCAGATGGGCATGGCGATCGCGCCGTTCGATAAACCGCAGCCG GTTCTGCGTGTTCAGCCGCTGGTTATCCCTGCTAGCGGTAAACTGCTGGATACCCTGGCTGC GTTGCCGGCTTTACCATCTTTAACCGGCCTGACTCAGCGTCAGCTGCAGCTGTCTATGGACC CGATGCTGGATCGTATGGCTTGCCGTCGTTAAgaattc
MGLAGLVLIEDEESGRLLLPKQWGIDDVPVIVQDKKFTAAGEIDYQLDVMSAAVGWFGDTLL TNGALYPEHAAPRGWLRLRLLNGCNARSLNFATSDKRPLYVVASDGGLLAEPVKVDELPVLM GERFEVLVDTSDGKPFDLVTLPVSQMGMAIAPFDKPQPVLRVQPLVIPASGKLLDTLAALPALP SLTGLTQRQLQLSMDPMLDRMACRR*
A0A377TQ76
ggatccATGGGCCTGGCGGGCCTGGTTCTGATCGAAGATGAAGAAAGCGGTCGTCTGCTGCTG CCGAAACAGTGGGGCATCGATGATGTTCCGGTGATCGTTCAGGATAAAAAATTCACCGCTG CGGGCGAAATTGACTACCAGCTGGATGTGATGTCTGCGGCGGTTGGTTGGTTCGGTGATAC CCTGCTGACCAACGGCGCACTGTACCCGGAACACGCGGCGCCGCGTGGCTGGCTGCGCCT GCGCCTGCTGAACGGCTGCAACGCGCGTTCTCTGAACTTCGCTACCAGCGATAAACGTCCG CTGTACGTTGTGGCTAGCGATGGCGGCCTGCTGGCTGAACCGGTGAAAGTTGATGAACTGC CGGTTCTGATGGGCGAACGTTTCGAAGTTCTGGTTGATACCTCTGATGGCAAACCGTTCGAT CTGGTTACTCTGCCGGTTAGCCAGATGGGTATGGCGATCGCACCGTTCGATAAACCGCAGC CGGTTCTGCGTGTTCAGCCGCTGGTTATCCCGGCTAGCGGCAAACTGCTGGATACCCTGGG CCGTAGTCCGGCGCTGCCGAGCCTGACCGGTCTCACCCAGCGGCAGCTGCAGCTGAGCAT GGGCTCTGATGCGCGTCCGGATGGTCACGCGGGCGCTGATGGTGAAGTTTGGCGTCCGGGC GATGGCCGTAACGGCAGTCGTCACGACGGTGCCTGGCGCCACGAACGTCATGGCGAATATG CTTCTCGTCGTCATGAACACGAACCGCGTATTCGTCATTGGTCTATGGCTTGTCGTCAAGCA CGTGCACTGATTAGCATTACCCCGACCGCTAGCACCGCTAAACCGAGCACCTAAgaattc
MGLAGLVLIEDEESGRLLLPKQWGIDDVPVIVQDKKFTAAGEIDYQLDVMSAAVGWFGDTLL TNGALYPEHAAPRGWLRLRLLNGCNARSLNFATSDKRPLYVVASDGGLLAEPVKVDELPVLM GERFEVLVDTSDGKPFDLVTLPVSQMGMAIAPFDKPQPVLRVQPLVIPASGKLLDTLGRSPALPS LTGLTQRQLQLSMGSDARPDGHAGADGEVWRPGDGRNGSRHDGAWRHERHGEYASRRHEH EPRIRHWSMACRQARALISITPTASTAKPST*
A0A1M4LFV1
ggatccGTGCTGGTTGATTTTAGCAACGGCGAAGCTGTTGATCTGGTTACCTATGGTGATAATG GTAGCGGTGATGGCCTGCATCTGATGCGTTTCGCGGTTGATCCGGCATTAGAAGGTCGTGTT GCTAAACAGCCTGAATCTTTAGATGGTCCGGCGGCTCCGGATGAAAAACTGTCTGTTCAGC GTCGTAGCTTTTTCTTTGATGAACGCATGGCAGAAAACATGAAACTGATGATGCGTCAGCC GTCCTCTAACCCGCACGCATCTGGTGATGATATGGATCACATGGAAATGGGCTCTATGGCGG GTATGGATCATGATATGCATGGTAGCCGTAGCGCGGCTGATGCTGGTCCAGCACTGGATGCA CTGACCTCTGGTGTTCAGATGGCGATTGCGGATAAACCGTTCGACATGGAACGTATCGATGT TGAAGCTAAATTGGGTTCTTGGGAAATCTGGGAACTGACTTCCCGTGAAATGGCTCATCCG TTCCATATCCATGGTGCGTCTTTCCGTATTCTGTCTATGAACGGTAAAAAACCGCCGGCGCA TCAGGCTGGCTGGAAAGATACCGCACTGATCGATGGTAAAGCAGAAATCCTGGTTCACTTC GATCGTGAAGCGGCGCGTTCTCACCCGTTCATGTTCCACTGCCATCTGCTGGAACATGAAG ATGTTGGTATGATGGCGCAGTTCGTTACCGTTTAAgaattc

VLVDFSNGEAVDLVTYGDNGSGDGLHLMRFAVDPALEGRVAKQPESLDGPAAPDEKLSVQRRS
FFFDERMAENMKLMMRQPSSNPHASGDDMDHMEMGSMAGMDHDMHGSRSAADAGPALD
ALTSGVQMAIADKPFDMERIDVEAKLGSWEIWELTSREMAHPFHIHGASFRILSMNGKKPPAH
QAGWKDTALIDGKAEILVHFDREAARSHPFMFHCHLLEHEDVGMMAQFVTV

*All sequences were codon optimized and the signal peptides were removed.

The structural characterization of the top-performing MCOs reveals unique features

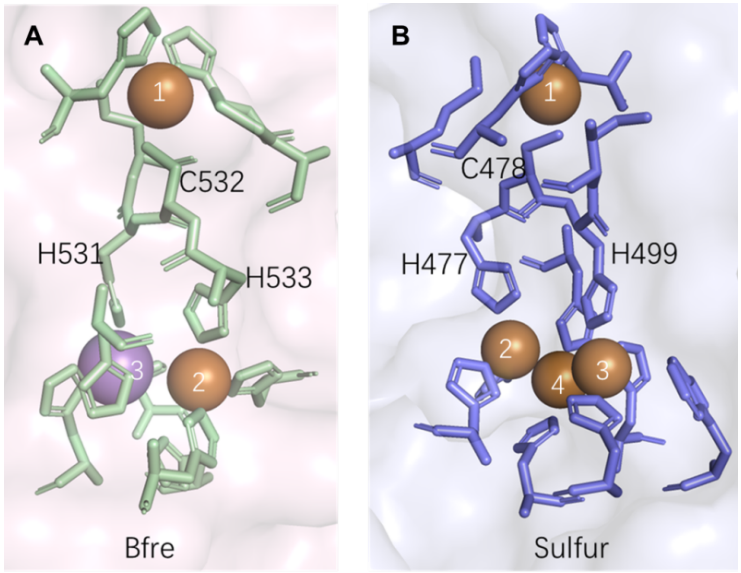


Figure S6 The amino acids coordinated with metal atoms in the active center.

A. The coordinated amino acids of Bfre are listed as follows:

- Cu1: H459, C532, H537, M542
- Cu2: H531, H202, H464
- Mn3: H533, H154, H200

B. The coordinated amino acids of Sulfur are listed as follows:

- Cu1: H424, M488, C478, H483
- Cu2: His477, His166, His429
- Cu3: His164, His126, His479
- Cu4: His124, His427

Table S9 Refinement statistics obtained for the final Bfre model

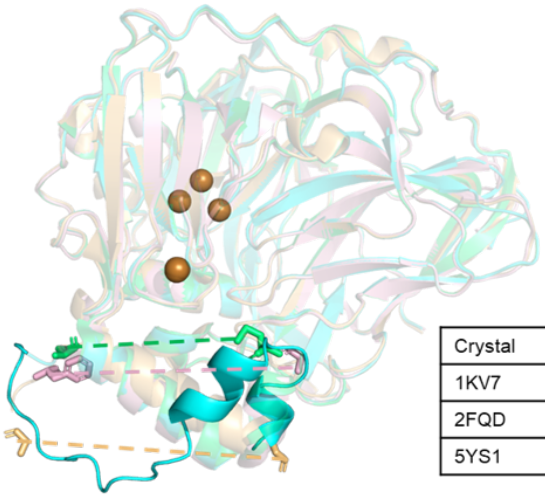
Data set	**Bfre**
PDB code	8Z5B
Data collection	
Wavelength (Å)	0.979
Resolution (Å)	29.02 - 1.4 (1.45 - 1.4)
Space group	P 4_12_12
Cell dimensions	74.59 74.59 230.946 90 90 90
Unique reflections	124241 (10796)
Completeness (%)	96.30 (85.04)
Rmeas (%)	3.0 (38.0)
CC1/2 (%)	99.8 (65.8)
Redundancy	14.3 (10.2)
$I/\sigma(I)$	24.2(1.2)
Refinement	
Rwork (%)	15.8 (2996)
Rfree (%)	17.3 (35.1)
R.m.s.d.	
Bond length (Å)	0.008
Bond angles (°)	1.04
No. of atoms	4834
protein	4143
water	688
Average B factors (Å^2)	23.14
protein	21.19
water	34.92
Ramachandran plot (%)	
Favored region	96.83
Allowed region	3.17
Outliers	0

Table S10 Refinement statistics obtained for the final Sulfur model

Data set	**Sulfur**
PDB code	8Z59
Data collection	
Wavelength (Å)	0.979
Resolution (Å)	50-2.58 (2.62-2.58)
Space group	P 1
Cell dimensions	86.6 103.4 131.9 104.9 104.0 98.4
Unique reflections	126901 (5355)
Completeness (%)	96.5 (81.1)
Rmeas (%)	14.9 (65.6)
CC1/2 (%)	98.4 (76.6)
Redundancy	3.4 (3.0)
$I/\sigma(I)$	7.4 (1.1)
Refinement	
Rwork (%)	19.7 (29.6)
Rfree (%)	23.7 (34.0)
R.m.s.d.	
Bond length (Å)	0.012
Bond angles (°)	1.47
No. of atoms	29113
protein	28636
water	445
Average B factors (Å^2)	50.75
protein	50.86
water	43.84
Ramachandran plot(%)	
Favored region	95.89
Allowed region	4.11
Outliers	0

Table S11 Comparison of protein secondary structures and residue interactions between Sulfur and Eclac

	Sulfur	Eclac
Salt bridge	72	44
Hydrogen Bond	368	368
Sheet	194	174
Helix	38	56
Coil	115	135



Crystal	AlphaFold	RMSD
1KV7	P36649	0.185
2FQD	P36649	0.244
5YS1	P36649	0.233

Figure S7 Superimposed alignments of the Eclac AlphaFold structure with its corresponding crystal structures. The structures are colored as follows: 1KV7 in green, 2FQD in pink, and 5YS1 in cyan. The dashed line indicates the regions that are missing in the structures. The relatively low RMSD values of the structural alignments indicate that the AlphaFold predictions are highly accurate.

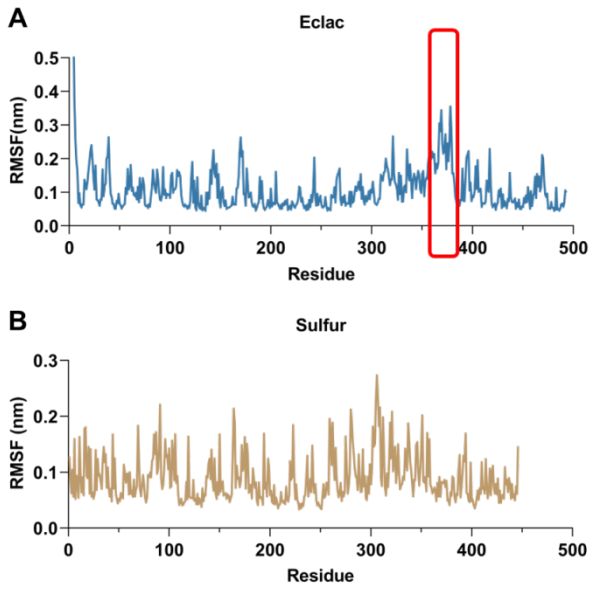


Figure S8 Root mean square fluctuation (RMSF) of Cα atoms per residue of (A) SulfurA and (B) Eclac at 37 °C. Source data are provided as a Source Data file.

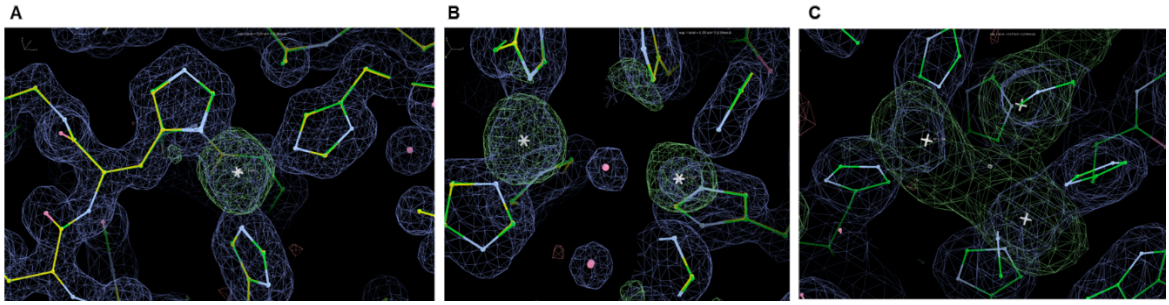


Figure S9 Fo-Fc omit and electron density maps of the crystal structures. A Mn atom in the Bfre crystal structure. **B** Cu atoms in the Bfre crystal structure. **C** Cu atoms in the Sulfur crystal structure.

Characterization of the L-Asparaginases

Development of a deep learning model to identify L-Asparaginases

To further verify the generalizability of the ESM-Ezy method, we manually collected a positive dataset of L-Asparaginases and sourced a negative dataset from Swiss-Prot, adhering to the same quality requirements as the MCO model. During the fine-tuning process of the L-Asparaginases model, we conducted five-fold cross-validation and employed the fine-tuned binary classification model to identify positive L-Asparaginases. Subsequently, we utilized the representations from the fine-tuned model to calculate the Euclidean distance between QEs and positive L-Asparaginase sequences, selecting the top-ranking candidates.

The expression and purification of the recombinant L-Asparaginases

The genes of the L-Asparaginases were synthesized by GeneScript (Nanjing, China) and cloned into plasmid pET28a. Then the strain *E.coli* BL21(DE3) was used for the expression of enzymes. Recombinant strains expressing the L-asparaginase were inoculated in 500 mL LB (kanamycin 50 mg/mL) broth and shaken at 37 ℃. The 0.1 mM IPTG was added into the culture when OD600 reached 0.6. The culture was shaken at 16 ℃ for another overnight for protein expression. Harvested cells were first centrifuged at 8000 rpm for 10 min and re-suspended in 50 mM phosphate buffered saline (PBS, pH 7.4) supplementing with 300 mM NaCl. The cells were disrupted using ultrasonication. After centrifuging at 12,000 rpm for 0.5 h, the supernatant was used as a crude enzyme mixture, which was further purified by nickel-affinity chromatography. Before sample loading, the 2-mL nickel-charged Ni-NTA agarose column (Qiagen, Shanghai, China) was equilibrated with binding buffer (50 mM PBS, 300 mM NaCl, pH 7.4), and elution buffer (50 mM PBS, 300 mM NaCl and 500 mM imidazole, pH 7.4) was used to elute enzyme preparations by gradient elution. Imidazole was removed by ultrafiltration. All of the enzyme purifications were carried out at low temperature.

Enzymatic activity assays

The Nessler's reagent (TMRM Co., Ltd., Beijing, China) method was adopted to determine L-asparaginase activity with modifications. The enzymatic reaction solution consisted of 0.1 -1 mM enzyme, 900 μL of 20 mM substrate L-asparagine of 50 mM Tris-HCl buffer (pH 7.4) and was incubated at 37 ℃ for 5 min. Add 100 μL of 25% (v/v) trichloroacetic acid (Macklin Co., Ltd., Shanghai, China) solution to terminate the enzymatic reaction. Then 10 μL of the above reaction supernatant coupled with 100 μL of Nessler's reagent were mixed with 890 μL of deionized water. After 10 min, ammonia release quantity was determined by OD_{436} at room temperature. One Unit of L-asparaginase activity was defined as the amount of enzyme releasing 1 μmol of ammonia per minute at pH 7.4 at 37 ℃.

Substrate specificity and kinetics parameters

The substrate specificity of the enzymes was determined through measuring the relative enzymatic activity toward L-asparagine , L-glutamine. The Km and Vmax of purified variants were determined by changing the concentration of L-asparagine over a range encompassing 0.5-5 mM at pH 7.4 at 37 ℃. LineweavereBurk plot was used to calculate the kinetics parameters K_m, k_{cat}

Table S12. The five-fold evaluation results of the L-Asparaginases model

Fold	Accuracy	Precision	Recall	F1-score	ROC-AUC
1	0.9971	0.9688	0.9394	0.9538	0.9883
2	0.9971	0.9412	0.9697	0.9552	0.9999
3	0.9990	0.9706	1.0000	0.9851	1.0000
4	0.9952	0.9118	0.9394	0.9254	0.9769
5	0.9961	1.0000	0.8788	0.9355	0.9890
Avg.	0.9969	0.9585	0.9455	0.9510	0.9908

Table S13 Characterization of the L-Asparaginases

Uniprot ID	Specific activity (U/mg)	Halt-life $t_{1/2}$ 50℃ (min)	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1}/mM)
O34482(query)	239.45±24.32	65.16±4.22	0.39±0.06	57.27±1.93	149.75±23.97
A0A3N5F6J4	469.96±58.86	23.06±1.45	1.09±0.19	112.91±8.49	105.752±12.29
A5EVZ9	262.71±16.46	42.91±4.31	0.63±0.08	82.01±3.31	131.86±18.64
H1D2G7	989.52±125.41	53.13±4.22	1.33±0.24	155.08±16.06	121.30±26.04
A0A2N0LBY7	30.06±3.09	34.06±3.54	ND	ND	ND
A0A085ZTY1	100.12±9.93	7.15±1.04	ND	ND	ND

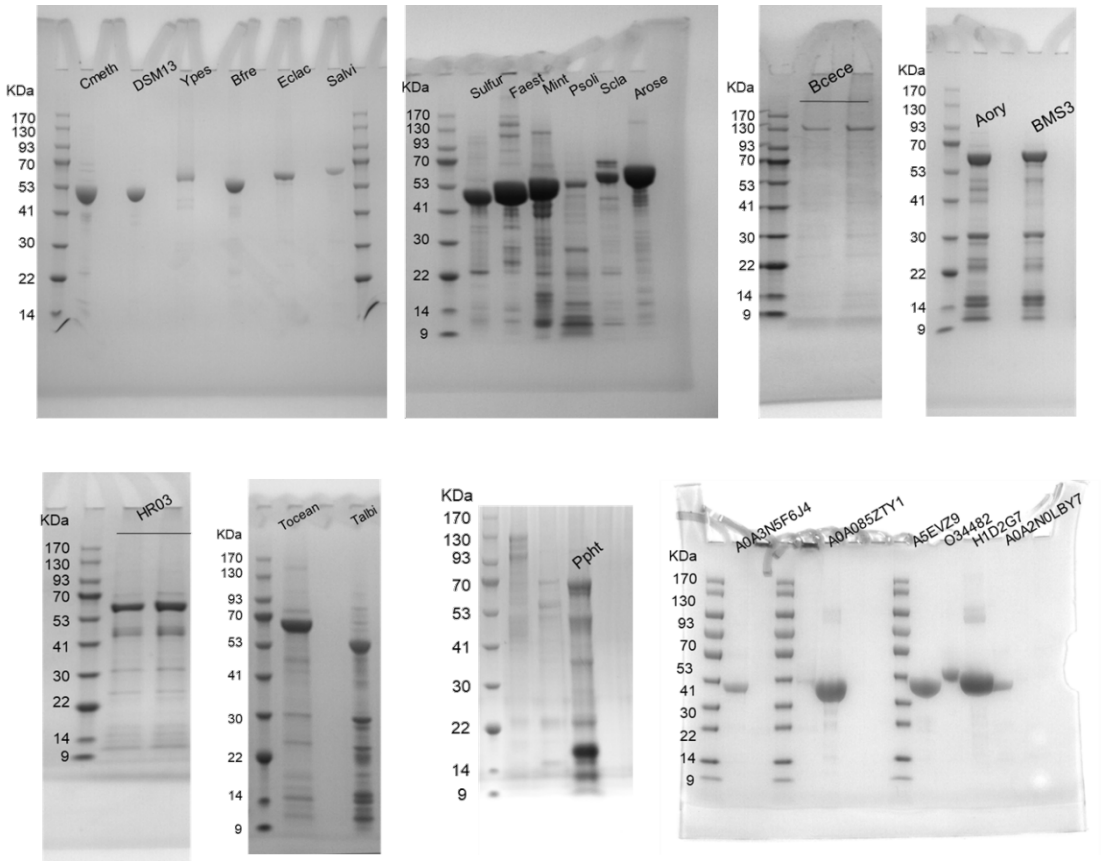


Figure S10 SDS-PAGE analysis of enzymes utilized in this study. The names of purified proteins are labeled on the gel.