



Mining and rational design of psychrophilic catalases using metagenomics and deep learning models

Shuning Wu^{1,2} · Guoshun Xu³ · Yongping Su^{1,2} · Huoqing Huang³ · Xinxin Xu¹ · Yuhong Zhang¹ · Jian Tian^{1,3} · Wei Zhang¹ · Zhiwei Zhang² · Bo Liu¹

Received: 29 July 2023 / Revised: 21 September 2023 / Accepted: 30 September 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

A complete catalase-encoding gene, designated *soiCat1*, was obtained from soil samples via metagenomic sequencing, assembly, and gene prediction. *soiCat1* showed 73% identity to a catalase-encoding gene of *Mucilaginibacter rubeus* strain P1, and the amino acid sequence of *soiCAT1* showed 99% similarity to the catalase of a psychrophilic bacterium, *Pedobacter cryoconitis*. *soiCAT1* was identified as a psychrophilic enzyme due to the low optimum temperature predicted by the deep learning model Preoptem, which was subsequently validated through analysis of enzymatic properties. Experimental results showed that *soiCAT1* has a very narrow range of optimum temperature, with maximal specific activity occurring at the lowest test temperature (4 °C) and decreasing with increasing reaction temperature from 4 to 50 °C. To rationally design *soiCAT1* with an improved temperature range, *soiCAT1* was engineered through site-directed mutagenesis based on molecular evolution data analyzed through position-specific amino acid possibility calculation. Compared with the wild type, one mutant, *soiCAT1*^{S205K}, exhibited an extended range of optimum temperature ranging from 4 to 20 °C. The strategies used in this study may shed light on the mining of genes of interest and rational design of desirable proteins.

Key points

- Numerous putative catalases were mined from soil samples via metagenomics.
- A complete sequence encoding a psychrophilic catalase was obtained.
- A mutant psychrophilic catalase with an extended range of optimum temperature was engineered through site-directed mutagenesis.

Keywords Catalase · Optimum temperature · Metagenomics · Deep learning model · Rational design

Introduction

Catalase (EC 1.11.1.6), which is present in all aerobic organisms and catalyzes hydrogen peroxide into water and oxygen, acts as an oxidoreductase enzyme to eliminate reactive oxygen species (ROS) produced through intracellular metabolism of molecular oxygen (Kaushal et al. 2018; Yamamoto et al. 2019). Aside from resistance to oxidative stress (Pradhan et al. 2017), catalase is also involved in responses to other environmental stresses including low temperature (Zhang et al. 2021), drought (Ma et al. 2017), and salt (Gondim et al. 2012). Extracellularly, catalases are effective candidates for eliminating residual H₂O₂ during many production processes, and therefore are used in various industrial fields including bleaching, food preservation, bioremediation, and wastewater treatment (Jia et al. 2016; Kaushal et al. 2018). Additionally, various industries employ

Shuning Wu and Guoshun Xu contributed equally.

✉ Zhiwei Zhang
zhiweizhang2012@163.com

✉ Bo Liu
liubo01@caas.cn

¹ Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, No.12 Zhongguancun South St., Haidian District, Beijing 100081, China

² College of Forestry, Shanxi Agricultural University, No.81 Longcheng Street, Taiyuan 030031, Shanxi, China

³ Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, No.2 Yuanmingyuan West Road, Haidian, Beijing 100193, China

catalases with specific enzymatic properties, such as thermostable catalases in textile bleaching and papermaking (Ebara and Shigemori 2008; Paar et al. 2001).

Low temperature helps to reduce microbial contamination and raw material deterioration during food processing, and some psychrophilic enzymes are widely used in food production as cost-effective and environmentally friendly additives. In contrast to the large amount of research available on alkali-stable and thermostable catalases (Calandrelli et al. 2008; Fall et al. 2023; Fu et al. 2014; Paar et al. 2001; Shaeer et al. 2019; Thompson et al. 2003), few psychrophilic catalases have been reported, including a catalase from *Antarctic Bacillus* with an optimum temperature of 25 °C (Wang et al. 2008) and a catalase from *Vibrio rumoiensis* sp. nov. with an optimum temperature of 30 °C (Yumoto et al. 2000). Demand for psychrophilic catalases with high catalytic activity and good stability at low temperatures is extensive.

Traditionally, the genes encoding catalases in various organisms have been identified via polymerase chain reaction (PCR), which is a time- and effort-consuming process that involves the isolation, enrichment culturing and identification of strains, extraction of chromosomal DNA, and PCR amplification of the target gene. More importantly, numerous microbes known as unculturable microbes cannot be cultured under artificial culture conditions, and thus their tremendous genetic resources cannot be accessed via the classical procedures of gene acquisition. The multitudinous valuable genes harbored in unculturable microbes (Datta et al. 2020) have remained largely mysterious (Wu et al. 2020). Fortunately, with the development of methods independent of microbial cultivation such as metagenomics based on DNA sequencing technology, genes of interest can be efficiently mined from samples containing unculturable microbes (Berini et al. 2017; Ngara and Zhang 2018).

Soil is a diverse and complex ecosystem (Jansson and Hofmockel 2020) that functions as a huge pool of microbes (Banerjee and van der Heijden 2023), biochemical gene library (Daniel 2005), and protein bank (Galhardi et al. 2020). Large numbers of genes identified from soil samples containing diverse unculturable microorganisms have been explored (Xu et al. 2022), revealing the microbial diversity of soil samples. Psychrophiles are diverse and widely distributed on Earth (Piette et al. 2011), and antioxidative enzymes such as catalases and superoxide dismutases are abundant in psychrophiles due to the increased solubility of gases such as O₂ at low temperatures, which allows for more ROS production during normal metabolism. Thus, soil samples from extremely cold and high-elevation areas serve as excellent pools for mining of genes encoding psychrophilic catalases.

Many natural proteins are unsuitable for industrial application due to low thermostability, narrow optimum

temperature or pH ranges, or low catalytic activity, and such proteins can be redesigned via protein engineering to improve their properties. In recent decades, error-prone PCR coupled with direct screening or selection was commonly employed to engineer proteins. This process was not dependent on protein structure but required a large amount of effort for the screening of numerous mutants. With the growth of databases containing protein sequence, structure, and function data and rapidly emerging applications of deep learning algorithms, site-directed mutagenesis mediated by rational design is playing increasing roles in protein engineering (Yang et al. 2019) and has been used to improve biochemical characteristics such as thermostability (Yoshida et al. 2021) and alkaline stability (Suplatov et al. 2014) as well as to design new proteins (Callaway 2022). Overall, protein engineering guided by machine-learning models can greatly promote the progress of protein evolution.

In this study, genes encoding psychrophilic catalases were identified from high-elevation soil samples through the combination of metagenomics with the optimum temperature prediction software Preoptem, and then a psychrophilic catalase encoded by the gene designated *soiCat1* was characterized. The optimum temperature range of *soiCAT1* was extended through site-directed mutagenesis guided by position-specific amino acid probabilities (PSAP) calculation.

Materials and methods

Mining of psychrophilic catalases via the deep learning model Preoptem

Soil sample collection and soil metagenomic analysis of Tianshan No. 1 Glacier was previously performed (Xu et al. 2022) and the data was deposited with the National Center for Biotechnology Information (NCBI) as BioProject PRJNA658179. The genes for catalases were identified via retrieval of DNA sequences from the NCBI database. The optimum temperatures of catalases were predicted using the Preoptem model, which was previously developed in our lab (<http://www.elabcaas.cn/pird/preoptem.html>).

Expression and purification of catalases

The pET-28a(+) vector was stored in our laboratory. *Escherichia coli* Top10, used as the cloning host, and *E. coli* BL21, used as the expression host, were obtained from TransGen Biotech Co., Ltd. (Beijing, China). The catalase-encoding genes were synthesized by GenScript Biotech Co. (Nanjing, China), and ligated into pET-28a(+) to generate the plasmid pET-28a/cat. pET-28a/cat was transformed into *E. coli* BL21 to develop the catalase-expressing strain. Expression and purification of catalases

were routinely performed according to the pET system manual with the induction temperature at 16 °C and purification temperature at 4 °C. Finally, the purified proteins were assayed via 12% polyacrylamide gel electrophoresis (SDS-PAGE).

Catalase activity assay

Catalase activity was determined as described previously (Martins and English 2014) with some modifications. Briefly, 225 µl of 50 mM NaH₂PO₄–Na₂HPO₄ buffer (pH 7.0) containing H₂O₂ at a final concentration of 30 mM was added as the substrate to a 1.5 ml tube and preheated at 20 °C for 3 min, followed by the addition of 25 µl of enzyme solution and incubation at 20 °C for 3 min. The reaction was terminated through addition of 250 µl sulfuric acid (2 M). The absorbance of the reaction solution was measured at 240 nm. Enzyme activity (U) was defined as the amount of enzyme required to decompose 1 µmol of hydrogen peroxide in 1 min.

Characterization of recombinant catalases

The optimum temperature for catalase activity was measured across the range of 4–50 °C in 50 mM NaH₂PO₄–Na₂HPO₄ buffer (pH 7.0). Enzyme activity was also determined at pH 5.0 to 10.0 to identify the optimum pH for enzyme activity. The pH stabilities of catalases were investigated using the residual activities after incubation in buffers with various pH values for 1 h. Enzyme thermostability was assessed by measuring the residual activity after incubation at various temperatures and durations.

Parameters including K_m , V_{max} , and K_{cat} were acquired after fitting the data to the Michaelis–Menten equation with the software GraphPad Prism (London, UK).

Site-directed mutagenesis of wild-type catalase based on rational design

Proteins with amino acid sequences similar to soiCAT were collected from Uniparc database. The optimum temperatures of these proteins were first predicted using Preoptem, from which the proteins were classified into two catalogues, one with optimum temperatures above 50 °C and the other with optimum temperatures below 50 °C. The amino acid frequency for each site of soiCAT was analyzed using the Parepro program (<http://www.elabcaas.cn/pird/premuse.html>), and the amino acids at specific sites of soiCAT were mutated into amino acids found only in catalases with optimum temperatures above 50 °C.

Results

Psychrophilic catalases mined via metagenomics and a deep learning model

Numerous microbes survive in the low-temperature environment of glaciers, many of which remain uncultured under laboratory conditions. The genes of interest harbored in unculturable microorganisms cannot be cloned via the classical cultivation-dependent process. Due to their independence on culturing, metagenomic sequencing and genome assembly are useful strategies for investigating the overall number of microbial species and genetic diversity in glaciers (Liu et al. 2022).

In total, 161 genes encoding catalases were predicted using metagenomic data from a glacier, and the optimum temperatures of these catalases were predicted using the established online deep learning model Preoptem (<http://www.elabcaas.cn/pird/preoptem.html>). The temperature of an enzyme is positively correlated to the optimum temperature determined in an experimental test (Zhang et al. 2022). The predicted optimum temperature was corrected to match the actual optimum temperature determined in an experimental test using a linear equation (Zhang et al. 2022) (Fig. S1). Subsequently, 16 genes encoding catalases with the corrected optimum temperatures below 20 °C are listed in Table 1.

The integrity of these 16 genes was analyzed via comparison of the proteins encoded by them with a protein database, and most of the genes were found to be incomplete (Table 2). The catalase encoded by the gene PI-H_1, designated *soiCat1*, had the lowest optimum temperature, and its gene sequence is complete (Table 2 and S1). Sequence alignment indicated that *soiCat1* shows 73% identity to a catalase-encoding gene from *Mucilaginibacter rubeus* strain P1, while *soiCAT1* shows 96.8% identity and 99% similarity in amino acid sequence to a catalase from *Pedobacter cryoconitis* (GenBank accession number WP_183867858.1).

To further elucidate the phylogenetic relationships between *soiCAT1* and psychrophilic catalases of various species, phylogenetic analysis was conducted. From the BRENDA Enzyme Database, 7 catalases with optimum temperatures below 25 °C were obtained; meanwhile, 64 catalases were acquired by searching the protein database of NCBI using psychrophilic catalase as the query. Additionally, the top 54 proteins homologous to *soiCAT1* were identified using the Basic Local Alignment Search Tool for Proteins (BLASTP) with the sequence of *soiCAT1* as the query. Of these proteins, 111 remained after running the clustering tool CD-HIT with a similarity threshold of 0.9, and a phylogenetic tree of low-temperature peroxidases

Table 1 The predicted and corrected optimum temperatures of catalases by Preoptem

Gene Id	Predicted temperature (°C)	Corrected temperature (°C)	Gene Id	Predicted temperature (°C)	Corrected temperature (°C)
PI-H_1	26.51	8.53	PI-L_4	29.4	17.18
PI-H_2	26.79	9.36	PI-H_4	29.48	17.42
PI-L_1	27.03	10.08	PI-H_5	29.57	17.69
PI-M_1	27.08	10.23	PI-M_3	29.59	17.75
PI-H_3	28.75	15.23	PI-H_6	29.61	17.81
PI-L_2	28.99	15.95	PI-H_7	29.62	17.84
PI-L_3	29.07	16.19	PI-L_5	29.7	18.08
PI-M_2	29.34	17.00	PI-H_8	29.87	18.59

PI protein Id, H high altitude soil samples, M medium altitude soil samples, L low altitude soil samples

Table 2 Sequence integrity analysis of catalase-encoding genes by sequence alignment

Predicted class	Sequence length	e-value	Identity	Accession number (ACC)	ACC sequence length
PI-H_1	502	0	0.9701	WP_183867858.1	502
PI-H_2	317	0	0.8707	WP_183888344.1	715
PI-L_1	484	0	0.8802	MBA2547478.1	835
PI-M_1	486	0	0.7479	MBA3357709.1	755
PI-H_3	367	0	0.8704	WP_079688601.1	747
PI-L_2	394	0	0.7513	PMR82282.1	715
PI-L_3	220	9e ⁻¹²⁹	0.8318	NIT41933.1	338
PI-M_2	293	9e ⁻¹⁴⁶	0.7832	MBA3257592.1	607
PI-L_4	433	0	0.8681	AMY08171.1	792
PI-H_4	420	0	0.8964	WP_068400485.1	715
PI-H_5	390	0	0.8067	RYZ06154.1	807
PI-M_3	292	0	0.9178	MBA2705139.1	755
PI-H_6	400	0	0.8625	WP_200604791.1	501
PI-H_7	652	0	0.8009	MBC7989628.1	800
PI-L_5	292	0	0.9724	RPH75229.1	290
PI-H_8	480	0	0.7438	MBE0592941.1	714

was constructed (Fig. 1). The closest phylogenetic relationship of soiCAT1 was with the catalase of *Pedobacter* sp. Most members of genus *Pedobacter* are psychrophilic strains that grow at low temperatures of 1–25 °C (He et al. 2020; Margesin et al. 2003), and therefore soiCAT1 is most likely a psychrophilic catalase.

Characterization of soiCAT1

To confirm that soiCAT1 is a psychrophilic catalase with experimental data, soiCAT1 was expressed in *E. coli* and purified using Ni-NTA resin. The purity of proteins was checked using SDS-PAGE, and a single band corresponding to an expected molecular weight of approximately 68 kDa was observed (Fig. S2). Purified recombinant soiCAT1 showed optimal activity at the lowest test temperature (4 °C),

which markedly decreased with increasing temperature from 4 to 50 °C (Fig. 2A). Notably, the activity of soiCAT1 at 20 °C was 60% of that at 4 °C. The optimal pH for soiCAT1 activity was 9 (Fig. 2B). soiCAT1 exhibited excellent stability at low temperatures of 10 °C and 20 °C (Fig. 2C), but rapidly lost stability at temperatures above 30 °C (data not shown). Additionally, soiCAT1 exhibited better stability at pH 9 than other pH levels tested (Fig. 2D). These results suggest that soiCAT1 is indeed a psychrophilic catalase.

Engineering soiCAT1 through rational design

The narrow optimum temperature range of soiCAT1 indicates that a slight shift would significantly decrease its activity, which greatly limits its use in various industrial applications. Ideally, soiCAT1 should have high and

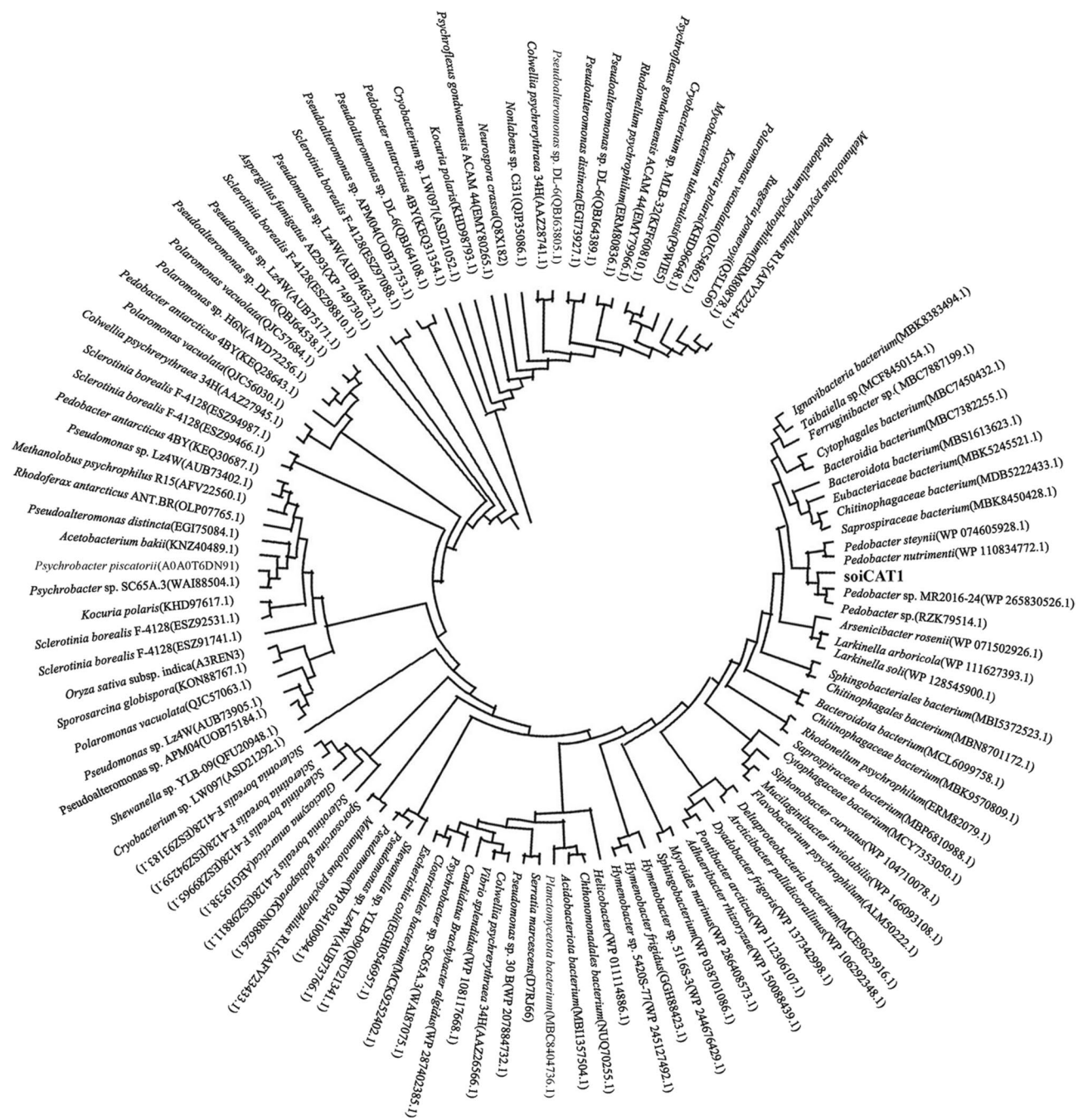


Fig. 1 Phylogenetic tree of catalase sequences using the Neighbor-Joining method in MEGA11

steady enzymatic activity across a wide temperature range. To broaden the optimum temperature range of *soiCAT1* and thus reduce its dependence on temperature, *soiCAT1* was rationally designed based on the coevolution of residues within protein sequences. First, over 14,000 proteins similar to *soiCAT1* were collected from Uniparc database. Preoptem was used to predict the optimum temperatures of the proteins, which were then divided into two groups: H and L. The H group contained 5565 proteins with optimum

temperatures above 50 °C (Table S2), while the L group comprised 8865 proteins with optimum temperatures below 50 °C (Table S3). Position-specific amino-acid frequencies (PSAPs) of the proteins in the two catalogues corresponding to each amino acid of *soiCAT1* were analyzed using the Parepro program (Meng et al. 2021; Tian et al. 2007) (Figs. S3 and S4), and the mutations of *soiCAT1* were screened to identify the amino acids that occur only with high optimum temperatures. Fifty-nine amino

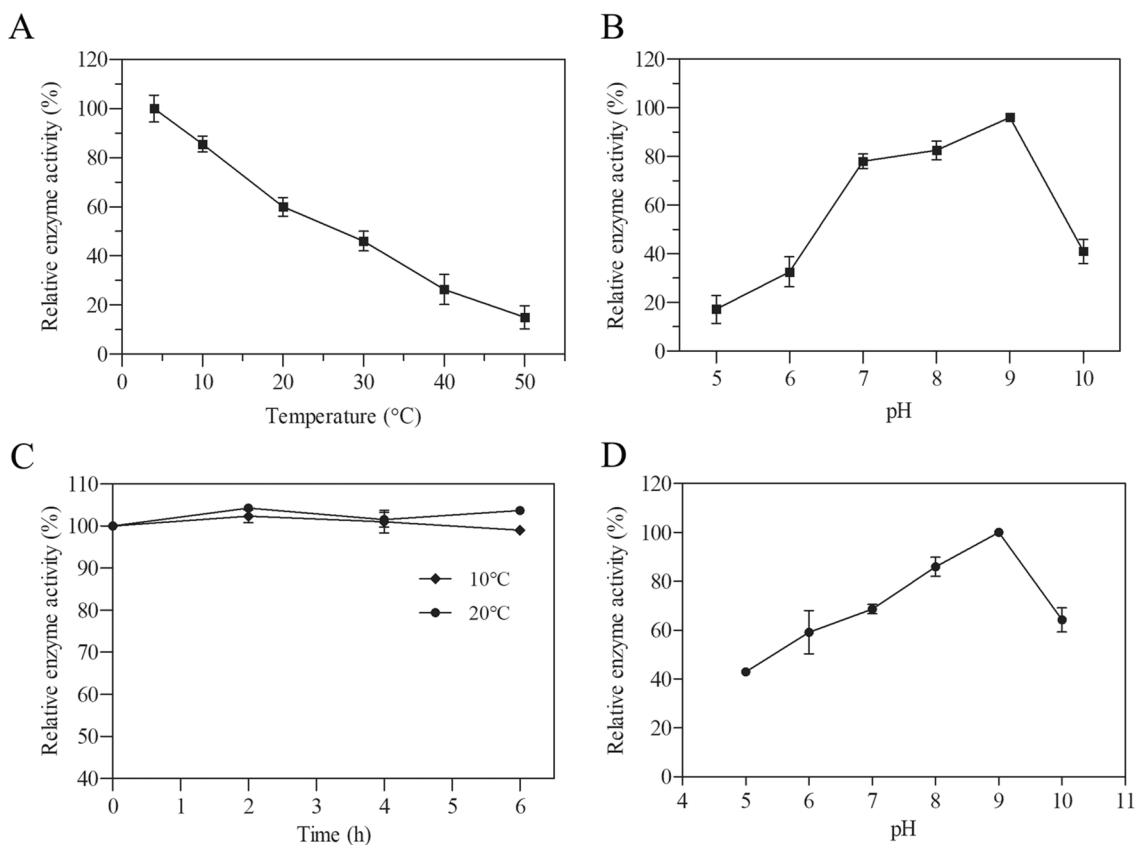


Fig. 2 Characterization of *soiCAT1*. **A** optimum temperature; **B** optimum pH; **C** thermal stability; **D** pH stability

Table 3 The amino acids of *soiCAT1* to be mutated to extend optimum temperature range

Ori AA	Site	CV	Mut AA
I	300	0.3292	M
T	63	0.2863	A
P	232	0.2804	E
N	446	0.2315	R
T	87	0.2153	P
K	264	0.1994	Y
H	42	0.1987	I
S	205	0.1863	K
E	110	0.169	V
E	296	0.1508	P
K	445	0.143	E
H	244	0.1403	Y
N	255	0.135	R
Q	223	0.1219	D

CV coefficient of variation

acid sites were found (Table S4), and the top 14 mutations were selected as candidates for the following experimental assessment (Table 3).

Properties of the mutant *soiCAT1*^{S205K}

Site-directed mutagenesis was employed to develop 14 mutants of *soiCAT1*, and the optimum temperatures of these mutants were determined after expression and purification. Among the 14 mutants, one mutant, *soiCAT1*^{S205K}, exhibited an extended optimum temperature range compared to *soiCAT1* and the other 13 mutants, and therefore *soiCAT1*^{S205K} was selected for further detailed characterization. *soiCAT1*^{S205K} showed high and steady enzymatic activity from 4 to 20 °C. *soiCAT1*^{S205K} had its highest activity at 4 °C, as did *soiCAT1*. However, the activity of *soiCAT1*^{S205K} remained steady over the range from 4 to 20 °C, in marked contrast to *soiCAT1*. Thereafter, the activity of *soiCAT1*^{S205K} decreased with increasing reaction temperature (Fig. 3A). *soiCAT1*^{S205K} activity showed similar dependence on pH, with an optimum pH of 9 (Fig. 3B). In terms of the effects of temperature and pH on catalase stability, *soiCAT1*^{S205K} exhibited similar characteristics to *soiCAT1* (Fig. 3C and D).

To examine the catalytic activity changes of *soiCAT1* and *soiCAT1*^{S205K} at different temperatures, the removal efficiency of H₂O₂ was assessed at 4 °C and 20 °C. At the same protein concentration (20 µg/L), nearly identical amounts of H₂O₂ were decomposed at 4 °C by *soiCAT1* and

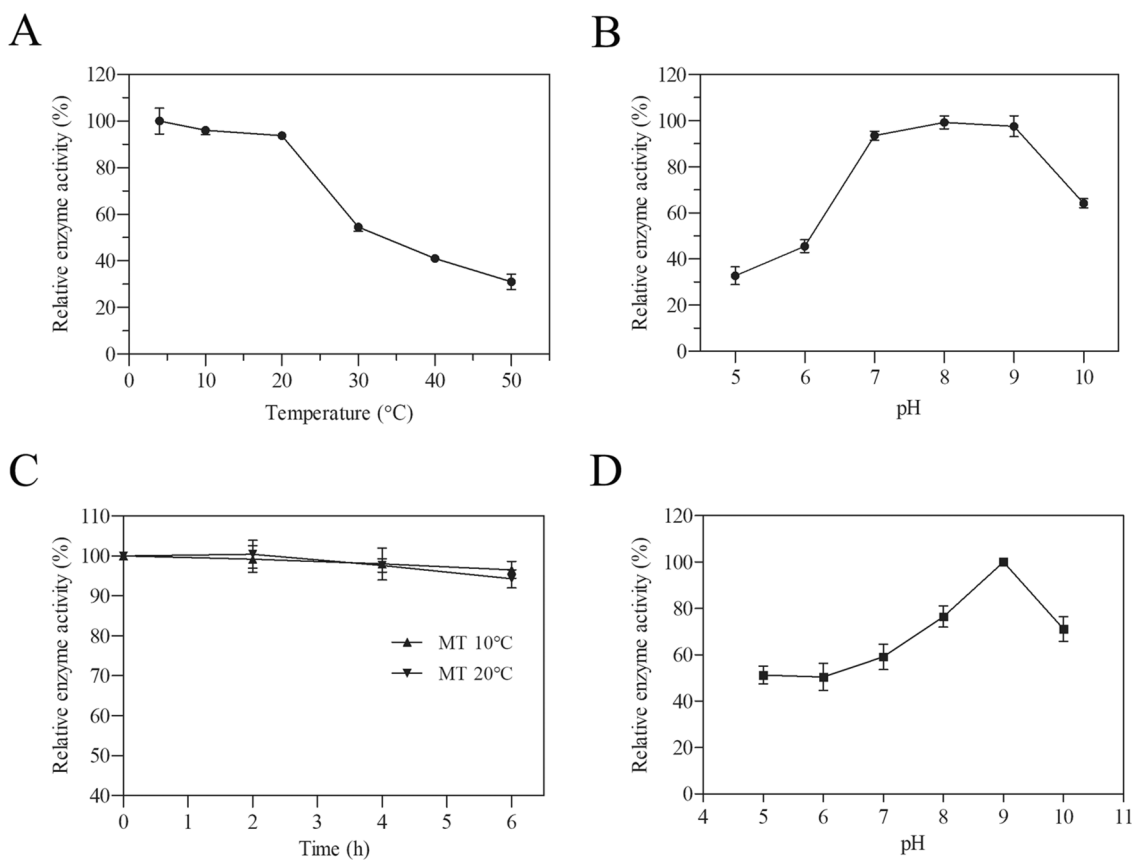
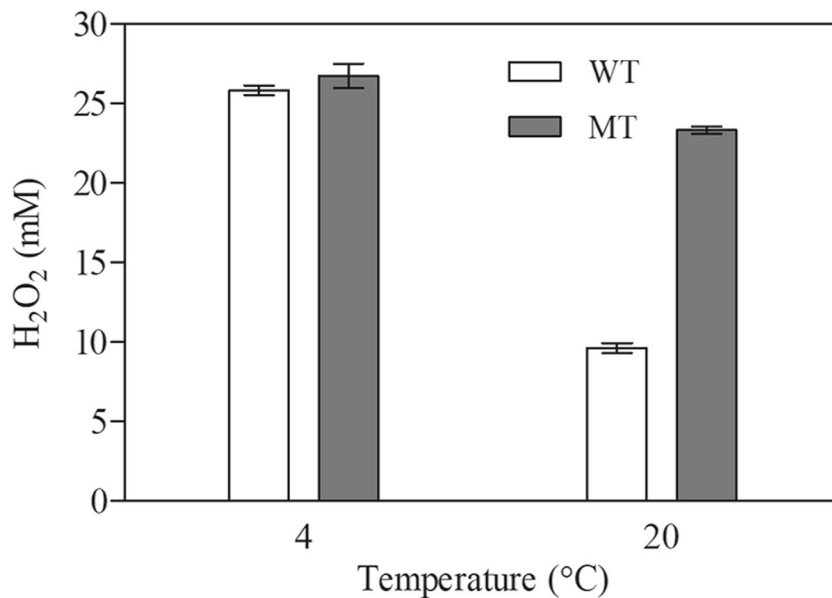


Fig. 3 Characterization of *soiCAT1*^{S205K}. **A** optimum temperature; **B** optimum pH; **C** thermal stability; **D** pH stability

soiCAT1^{S205K}. At 20 °C, the amount of H₂O₂ removed by *soiCAT1*^{S205K} was similar to that removed by *soiCAT1* or *soiCAT1*^{S205K} at 4 °C and markedly higher than the amount

removed by *soiCAT1* at 20 °C (Fig. 4). These results indicate that the temperature range over which *soiCAT1*^{S205K} maintains high catalytic activity was extended after site-directed

Fig. 4 The removal amounts of H₂O₂ of *soiCAT1* and *soiCAT1*^{S205K} at different temperatures



mutagenesis via rational design. Kinetic parameters of the wild type and mutant at 20 °C were also analyzed and fitted using nonlinear regression by GraphPad Prism 6 (Fig. 5). The K_m values of soiCAT1 and soiCAT1^{S205K} were 46.31 mM and 32.07 mM, while the V_{max} values of soiCAT1 and soiCAT1^{S205K} were 1.39×10^6 $\mu\text{mol}/\text{mg}/\text{min}$ and 1.54×10^6 $\mu\text{mol}/\text{mg}/\text{min}$, respectively. Overall, soiCAT1^{S205K} exhibited a wider range of optimum temperature than soiCAT1, without great alteration of other important properties.

Discussion

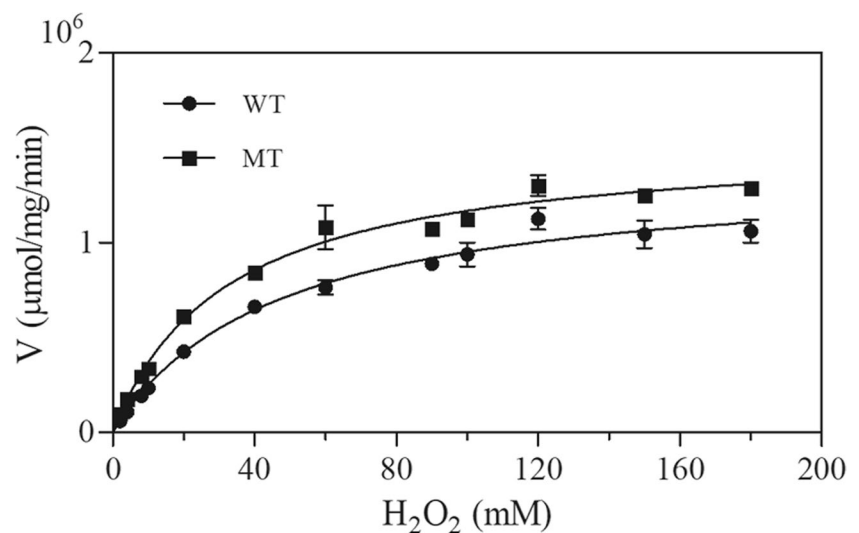
Psychrophilic catalases are considered effective candidates for food processing due to their high catalytic activity and stability at low temperature, which can strongly reduce the deterioration of raw materials as well as the risk of microbial contamination. Psychrophiles are the main providers of psychrophilic catalases. However, psychrophiles generally live under extremely cold conditions including glaciers, the deep sea and ice lakes, and most such microbes remain unculturable under artificial conditions, complicating the isolation of genes encoding psychrophilic catalases from these microbes. Fortunately, metagenomics methods are independent of culturing, providing a useful strategy for mining of microbial and genetic resources (Acinas et al. 2021; Daniel 2005; Xu et al. 2022), and a large number of genes have been mined via metagenomics. For these reasons, we aimed to identify psychrophilic catalase-encoding genes in a glacier sample from Tianshan, China, via metagenomics. The results suggest that this glacier harbors abundant catalase-encoding genes, and demonstrate that metagenomics is a useful high-throughput technique for identifying novel genes of interest from complex samples.

Efficient screening of desirable genes from enormous metagenomic datasets is another serious challenge. Optimum temperature is one of the key parameters characterizing an enzyme, and therefore accurate prediction of the optimum temperature of enzymes is useful for preliminary screening of desirable enzymes. We used the deep learning model Preoptem for high-throughput screening of potential psychrophilic catalases with optimum temperatures below 30 °C from among 161 catalases. soiCAT1 was predicted as a psychrophilic catalase via sequence alignment and this property was confirmed through enzymatic property analysis, validating the use of the deep learning model Preoptem as an effective tool to screen desired proteins from a large metagenomic dataset.

The catalytical activity of soiCAT1 peaked at 4 °C, and then decreased with increasing temperature, although it remained relatively stable at temperatures up to 20 °C. Notably, soiCAT1 had a narrow range of optimum temperature, which may limit its application. To broaden the optimum temperature range of soiCAT1, specific amino acids were mutated via PSAP to match the sequences of catalases with optimum temperatures above 50 °C. The mutant soiCAT1^{S205K} exhibited the expected property, with steady catalytical activity across the range of from 4 to 20 °C, confirming that site-directed mutagenesis via rational design is an efficient method for protein engineering.

In summary, we identified numerous catalase-encoding genes through analysis of soil metagenomic data and efficiently isolated putative psychrophilic catalases using the deep learning model Preoptem. One psychrophilic catalase was then engineered to improve specific properties through rational design guided by PASP calculation. The strategies used in this study are suitable for mining and rational design of other enzymes from gigantic metagenomic datasets.

Fig. 5 The kinetic curves of soiCAT1 and soiCAT1^{S205K} for H₂O₂ at 20 °C using nonlinear regression by GraphPad Prism 6.0



Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00253-023-12926-1>.

Author contribution Z.Z. and B.L. designed and supervised the research and analyzed the data. S.W., G.X., and Y.S. performed the experiments, analyzed the data, and wrote the manuscript. Q.H., X.X., Y.Z., J.T., and W.Z. analyzed the data and revised the manuscript.

Funding This work was supported by the National Key R&D Program of China (2021YFA0910602 and 2022YFA0912301) and Agricultural Science and Technology Innovation Program, CAAS-ZDRW202305; the China Agriculture Research System of MOF and MARA (CARS-41).

Data availability All materials are available by the corresponding author upon reasonable request.

Declarations

Ethical approval This work does not contain any studies with human participants or animals.

Conflict of interest The authors declare no competing interests.

References

- Acinas SG, Sanchez P, Salazar G, Cornejo-Castillo FM, Sebastian M, Logares R, Royo-Llonch M, Paoli L, Sunagawa S, Hingamp P, Ogata H, Lima-Mendez G, Roux S, Gonzalez JM, Arrieta JM, Alam IS, Kamau A, Bowler C, Raes J, Pesant S, Bork P, Agusti S, Gojobori T, Vaque D, Sullivan MB, Pedros-Alio C, Massana R, Duarte CM, Gasol JM (2021) Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities. *Commun Biol* 4:604. <https://doi.org/10.1038/s42003-021-02112-2>
- Banerjee S, van der Heijden MGA (2023) Soil microbiomes and one health. *Nat Rev Microbiol* 21:6–20. <https://doi.org/10.1038/s41579-022-00779-w>
- Berini F, Casciello C, Marcone GL, Marinelli F (2017) Metagenomics: novel enzymes from non-culturable microbes. *FEMS Microbiol Lett* 364:21. <https://doi.org/10.1093/femsle/fnx211>
- Calandrelli V, Gambacorta A, Romano I, Carratore V, Lama L (2008) A novel thermo-alkali stable catalase-peroxidase from *Oceanobacillus oncorhynchi* subsp. *incaldaniensis*: purification and characterization. *World J Microbiol Biotechnol* 24:2269–2275. <https://doi.org/10.1007/s11274-008-9741-3>
- Callaway E (2022) Scientists are using AI to dream up revolutionary new proteins. *Nature* 609:661–662. <https://doi.org/10.1038/d41586-022-02947-7>
- Daniel R (2005) The metagenomics of soil. *Nat Rev Microbiol* 3:470–478. <https://doi.org/10.1038/nrmicro1160>
- Datta S, Rajnish KN, Samuel MS, Pugazhendhi A, Selvarajan E (2020) Metagenomic applications in microbial diversity, bioremediation, pollution monitoring, enzyme and drug discovery. A review. *Environ Chem Lett* 18:1229–1241. <https://doi.org/10.1007/s10311-020-01010-z>
- Ebara S, Shigemori Y (2008) Alkali-tolerant high-activity catalase from a thermophilic bacterium and its overexpression in *Escherichia coli*. *Protein Expr Purif* 57:255–260. <https://doi.org/10.1016/j.pep.2007.09.015>
- Fall I, Czerwec Q, Abdellaoui S, Doumche B, Ochs M, Rmond C, Rakotoarivonina H (2023) A thermostable bacterial catalase-peroxidase oxidizes phenolic compounds derived from lignins. *Appl Microbiol Biotechnol* 107:201–217. <https://doi.org/10.1007/s00253-022-12263-9>
- Fu XH, Wang W, Hao JH, Zhu XL, Sun M (2014) Purification and characterization of catalase from marine bacterium *Acinetobacter* sp. YS0810. *Biomed Res Int* 2014:409626. <https://doi.org/10.1155/2014/409626>
- Galhardi JA, Fraceto LF, Wilkinson KJ, Ghoshal S (2020) Soil enzyme activities as an integral part of the environmental risk assessment of nanopesticides. *J Agric Food Chem* 68:8514–8516. <https://doi.org/10.1021/acs.jafc.0c04344>
- Gondim FA, Gomes E, Costa JH, Alencar NLM, Prisco JT (2012) Catalase plays a key role in salt stress acclimation induced by hydrogen peroxide pretreatment in maize. *Plant Physiol Biochem* 56:62–71. <https://doi.org/10.1016/j.plaphy.2012.04.012>
- He XY, Li N, Chen XL, Zhang YZ, Zhang XY, Song XY (2020) *Pedobacter indicus* sp. nov., isolated from deep-sea sediment. *Antonie Van Leeuwenhoek* 113:357–364. <https://doi.org/10.1007/s10482-019-01346-9>
- Jansson JK, Hofmockel KS (2020) Soil microbiomes and climate change. *Nat Rev Microbiol* 18:35–46. <https://doi.org/10.1038/s41579-019-0265-7>
- Jia X, Chen J, Lin C, Lin X (2016) Cloning, expression, and characterization of a novel thermophilic monofunctional catalase from *Geobacillus* sp. CHB1. *Biomed Res Int* 2016:7535604. <https://doi.org/10.1155/2016/7535604>
- Kaushal J, Mehandia S, Singh G, Raina A, Arya SK (2018) Catalase enzyme: application in bioremediation and food industry. *Bio-catal Agric Biotechnol* 16:192–199. <https://doi.org/10.1016/j.bcab.2018.07.035>
- Liu Y, Ji M, Yu T, Zaugg J, Anesio AM, Zhang Z, Hu S, Hugenholtz P, Liu K, Liu P, Chen Y, Luo Y, Ya T (2022) A genome and gene catalog of glacier microbiomes. *Nat Biotechnol* 40:1341–1348. <https://doi.org/10.1038/s41587-022-01367-2>
- Ma Y, Rajkumar M, Moreno A, Zhang C, Freitas H (2017) Serpentine endophytic bacterium *Pseudomonas azotoformans* ASS1 accelerates phytoremediation of soil metals under drought stress. *Chemosphere* 185:75–85. <https://doi.org/10.1016/j.chemosphere.2017.06.135>
- Margesin R, Sproer C, Schumann P, Schinner F (2003) *Pedobacter cryoconitis* sp. nov., a facultative psychrophile from alpine glacier cryoconite. *Int J Syst Evol Microbiol* 53:1291–1296. <https://doi.org/10.1099/ijs.0.02436-0>
- Martins D, English AM (2014) Catalase activity is stimulated by H₂O₂ in rich culture medium and is required for H₂O₂ resistance and adaptation in yeast. *Redox Biol* 2:308–313. <https://doi.org/10.1016/j.redox.2013.12.019>
- Meng X, Yang L, Liu H, Li Q, Xu G, Zhang Y, Guan F, Zhang Y, Zhang W, Wu N, Tian J (2021) Protein engineering of stable ISPETase for PET plastic degradation by Premuse. *Int J Biol Macromol* 180:667–676. <https://doi.org/10.1016/j.ijbiomac.2021.03.058>
- Ngara TR, Zhang H (2018) Recent advances in function-based metagenomic screening. *Genom Proteom Bioinf* 16:405–415. <https://doi.org/10.1016/j.gpb.2018.01.002>
- Paar A, Costa S, Tzanov T, Gudelj M, Robra KH, Cavaco-Paulo A, Gubitz GM (2001) Thermo-alkali-stable catalases from newly isolated *Bacillus* sp. for the treatment and recycling of textile bleaching effluents. *J Biotechnol* 89:147–153. [https://doi.org/10.1016/s0168-1656\(01\)00305-4](https://doi.org/10.1016/s0168-1656(01)00305-4)
- Piette F, D'Amico S, Mazzucchelli G, Danchin A, Leprince P, Feller G (2011) Life in the cold: a proteomic study of cold-repressed proteins in the antarctic bacterium *pseudoalteromonas haloplanktis* TAC125. *Appl Environ Microbiol* 77:3881–3883. <https://doi.org/10.1128/AEM.02757-10>
- Pradhan A, Herrero-De-Dios C, Belmonte R, Budge S, Garcia AL, Kolmogorova A, Lee KK, Martin BD, Ribeiro A, Bebes A, Yuecel

- R, Gow NAR, Munro CA, MacCallum DM, Quinn J, Brown AJP (2017) Elevated catalase expression in a fungal pathogen is a double-edged sword of iron. *PLoS Pathog* 13:e1006405. <https://doi.org/10.1371/journal.ppat.1006405>
- Shaer A, Aslam M, Rashid N (2019) A highly stable manganese catalase from *Geobacillus thermopakistanensis*: molecular cloning and characterization. *Extremophiles* 23:707–718. <https://doi.org/10.1007/s00792-019-01124-5>
- Suplatov D, Panin N, Kirilin E, Shcherbakova T, Kudryavtsev P, Svedas V (2014) Computational design of a pH stable enzyme: understanding molecular mechanism of penicillin acylase's adaptation to alkaline conditions. *PLoS One* 9:e100643. <https://doi.org/10.1371/journal.pone.0100643>
- Thompson VS, Schaller KD, Apel WA (2003) Purification and characterization of a novel thermo-alkali-stable catalase from *Thermus brockianus*. *Biotechnol Prog* 19:1292–1299. <https://doi.org/10.1021/bp034040t>
- Tian J, Wu N, Guo X, Guo J, Zhang J, Fan Y (2007) Predicting the phenotypic effects of non-synonymous single nucleotide polymorphisms based on support vector machines. *BMC Bioinformatics* 8:450. <https://doi.org/10.1186/1471-2105-8-450>
- Wang W, Sun M, Liu WS, Zhang B (2008) Purification and characterization of a psychrophilic catalase from Antarctic *Bacillus*. *Can J Microbiol* 54:823–828. <https://doi.org/10.1139/W08-066>
- Wu XQ, Spencer S, Gushgari-Doyle S, Yee MO, Voriskova J, Li YF, Alm EJ, Chakraborty R (2020) Culturing of “Unculturable” subsurface microbes: natural organic carbon source fuels the growth of diverse and distinct bacteria from groundwater. *Front Microbiol* 11:610001. <https://doi.org/10.3389/fmicb.2020.610001>
- Xu GS, Zhang LW, Liu XQ, Guan FF, Xu YQ, Yue HT, Huang JQ, Chen JY, Wu NF, Tian J (2022) Combined assembly of long and short sequencing reads improve the efficiency of exploring the soil metagenome. *BMC Genomics* 23:37. <https://doi.org/10.1186/s12864-021-08260-3>
- Yamamoto K, Uchida K, Furukawa A, Tamura T, Ishige Y, Negi M, Kobayashi D, Ito T, Kakegawa T, Hebisawa A, Awano N, Takemura T, Amano T, Akashi T, Eishi Y (2019) Catalase expression of *Propionibacterium acnes* may contribute to intracellular persistence of the bacterium in sinus macrophages of lymph nodes affected by sarcoidosis. *Immunol Res* 67:182–193. <https://doi.org/10.1007/s12026-019-09077-9>
- Yang KK, Wu Z, Arnold FH (2019) Machine-learning-guided directed evolution for protein engineering. *Nat Methods* 16:687–694. <https://doi.org/10.1038/s41592-019-0496-6>
- Yoshida K, Kawai S, Fujitani M, Koikeda S, Kato R, Ema T (2021) Enhancement of protein thermostability by three consecutive mutations using loop-walking method and machine learning. *Sci Rep* 11:11883. <https://doi.org/10.1038/s41598-021-91339-4>
- Yumoto I, Ichihashi D, Iwata H, Istokovics A, Ichise N, Matsuyama H, Okuyama H, Kawasaki K (2000) Purification and characterization of a catalase from the facultatively psychrophilic bacterium *Vibrio rumoiensis* S-1(T) exhibiting high catalase activity. *J Bacteriol* 182:1903–1909. <https://doi.org/10.1128/Jb.182.7.1903-1909.2000>
- Zhang H, Tan K, Zheng H (2021) Characterization of a novel catalase gene and its response to low temperature stress in noble scallop *Chlamys nobilis* with different total carotenoids content. *Aquac Res* 52:6179–6188. <https://doi.org/10.1111/are.15479>
- Zhang Y, Guan FF, Xu GS, Liu XQ, Zhang YH, Sun JL, Yao B, Huang HQ, Wu NF, Tian J (2022) A novel thermophilic chitinase directly mined from the marine metagenome using the deep learning tool Preoptem. *Bioresour Bioprocess* 9:54. <https://doi.org/10.1186/s40643-022-00543-1>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.