

Supplementary Materials for
**Anthropogenic activity and climate change exacerbate the spread of
pathogenic bacteria in the environment**

Yu Geng *et al.*

* Corresponding author. Email: liuya@sdu.edu.cn (Y.L.), zhangzheng@sdu.edu.cn (Z.Z.); ORCID:
0000-0002-1756-0907 (Y.L.), 0000-0001-9971-6006 (Z.Z.)

This PDF file includes:

Figs. S1 to S17
Table S6
Legends for Tables S1 to S5

Other Supplementary Materials for this manuscript include the following:

Tables S1 to S5

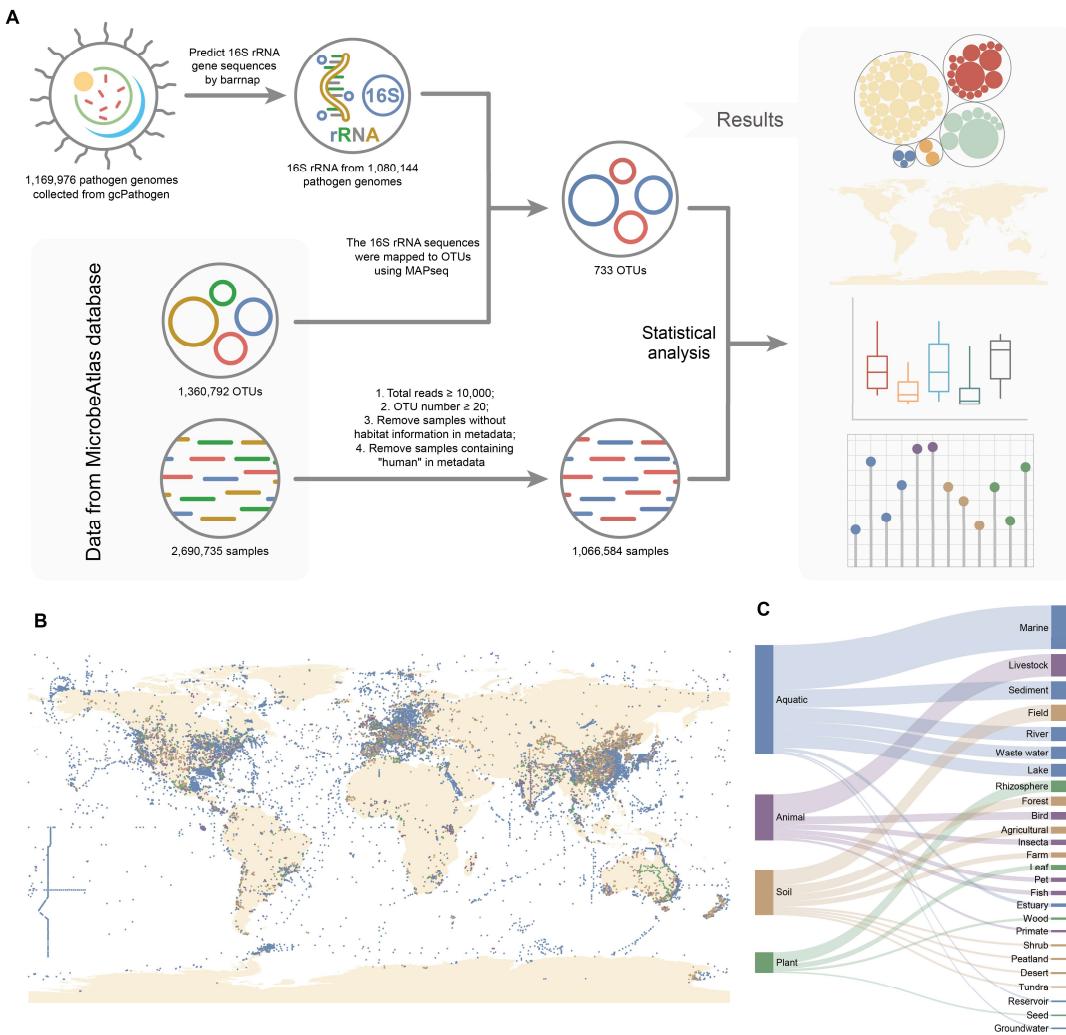


Fig. S1. Workflow of this study. (A) Identification process of HPB. (B) Geographical location of the samples. (C) Classification of habitats. Based on the metadata, the samples were classified into four categories: aquatic (blue), animal (purple), soil (brown), and plant (green). Additionally, these categories were further subdivided into 26 distinct microbial habitats.

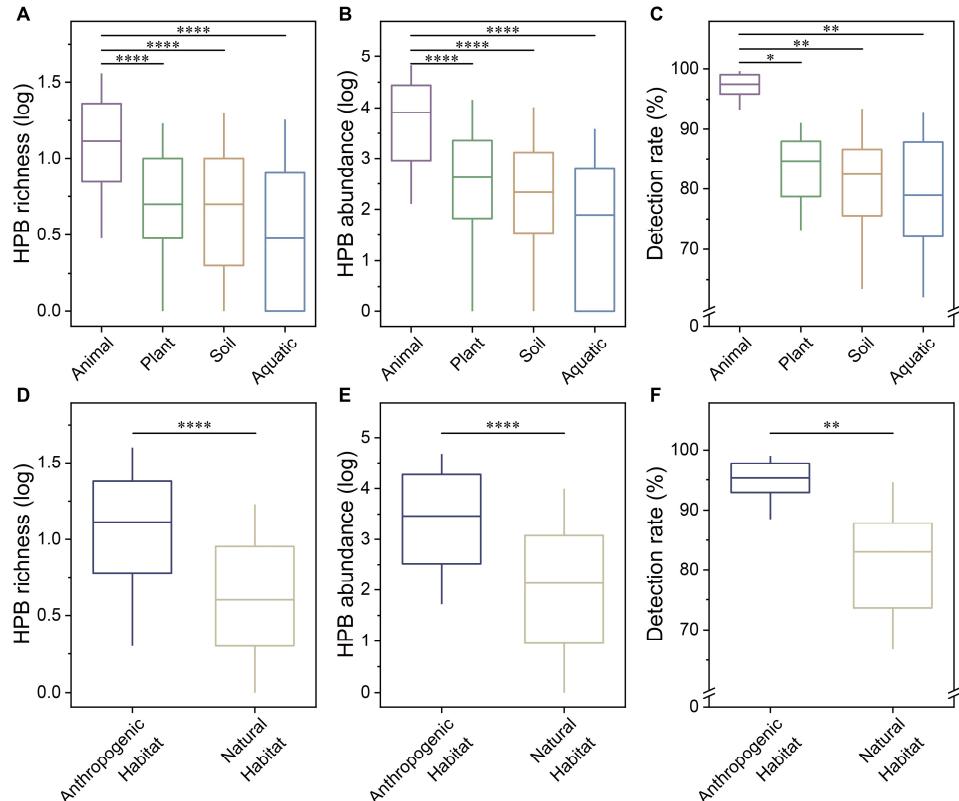


Fig. S2. HPB in different types of habitats. (A to C) Differences in HPB richness (A), abundance (B), and detection rates (C) across four habitats: animal, plant, soil, and aquatic. (D to F) Differences in HPB richness (D), abundance (E), and detection rates (F) between anthropogenic habitats and natural habitats. Comparisons between bins were conducted using the Wilcoxon rank-sum test, $*P < 0.05$, $**P < 0.01$, $****P < 0.0001$. In all the depicted boxplots, the middle line indicates the median, the box represents the 25th-75th percentiles, and the error bar indicates the 10th-90th percentiles of the observations. The richness and abundance values are log-transformed (base 10).

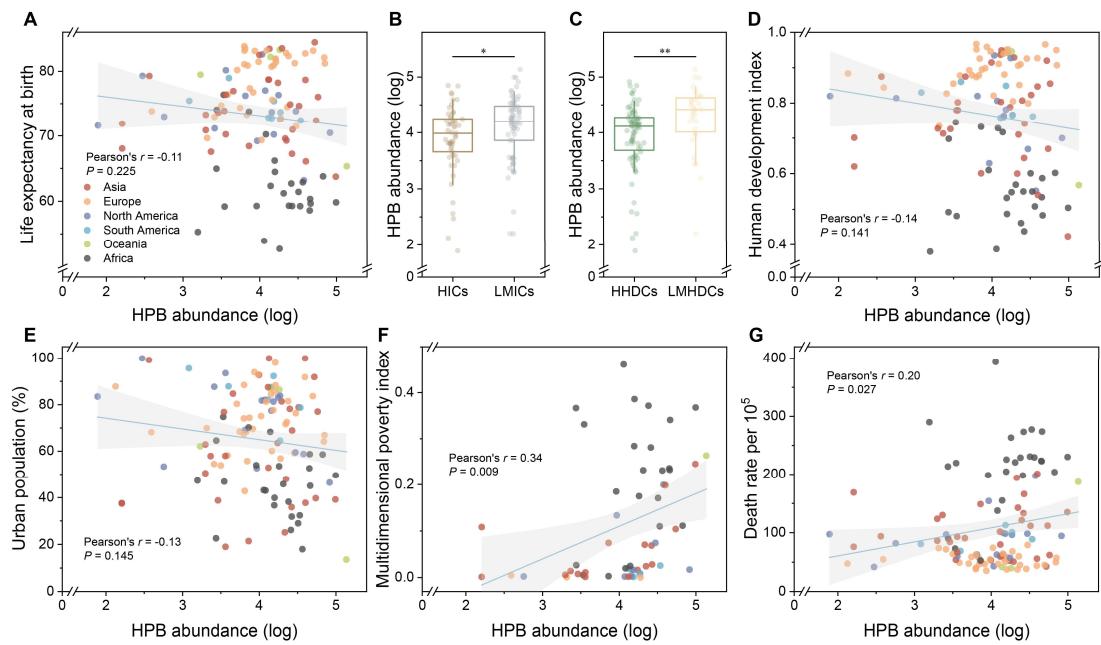


Fig. S3. Relationships between the abundance of HPB and anthropogenic activities. (A) Relationships between the abundance of HPB and life expectancy at birth (2022). Each dot represents a country, and the color indicates the continent where the country is located. Each country contains more than 30 samples. In all the depicted scatter plots, the lines indicate the best linear fit, and the shaded areas represent the 95% confidence intervals of the fitted curves. Pearson's correlation tests were used to examine the correlations, with P values indicating statistical significance. Abundance values are log-transformed (base 10). (B) Differences in the abundance of HPB among countries with different income levels (2024 fiscal year). High-income countries (HICs) are shown in brown, including "High income"; low- and middle-income countries (LMICs) are shown in gray, including "Upper middle income", "Lower middle income", and "Low income". (C) Differences in the abundance of HPB among countries with different levels of human development (2024). High human development countries (HHDCs) are shown in green, including "Very high human development" and "High human development"; low and middle human development countries (LMHDCs) are shown in light yellow, including "Medium human development" and "Low human development". Comparisons between bins were conducted using the Wilcoxon rank-sum test, * P < 0.05, ** P < 0.01. In all the depicted boxplots, the middle line indicates the median, the box represents the 25th-75th percentiles, and the error bar indicates the 10th-90th percentiles of the observations. (D to G) Relationships between the abundance of HPB and socioeconomic factors. The socioeconomic factors include Human Development Index (HDI, 2024, D), urban population (2022, E), Global Multidimensional Population Index (MPI, 2023, F), and mortality rate per 100,000 by pathogen (2019, G).

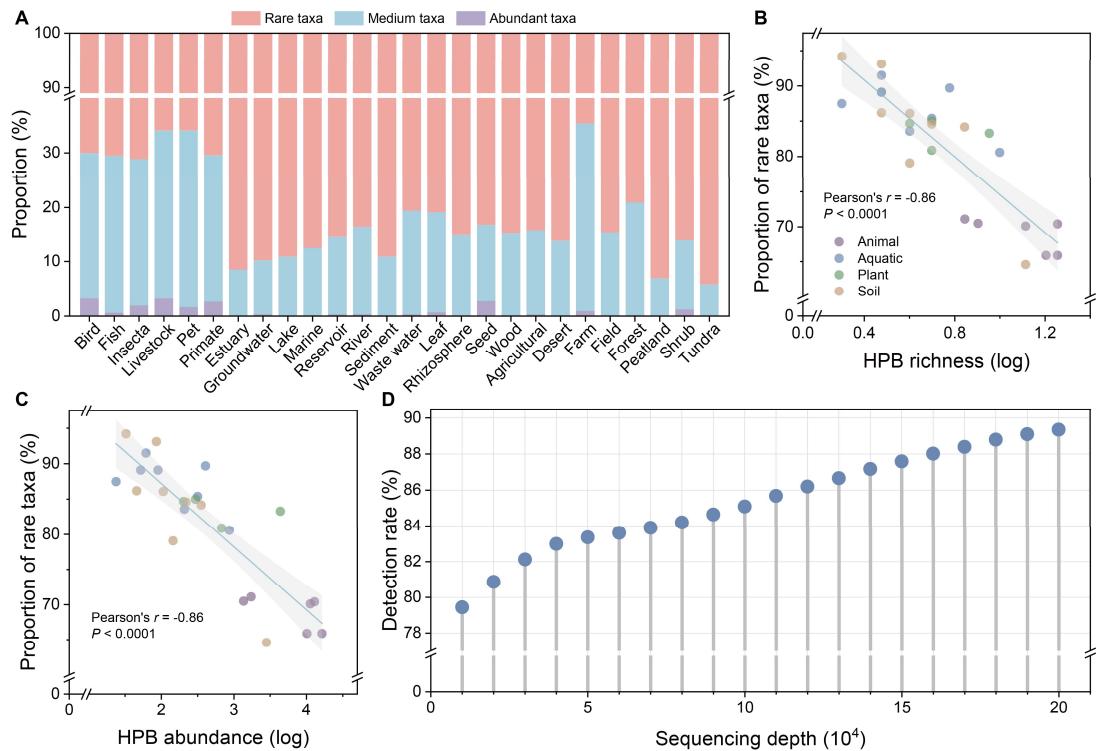


Fig. S4. HPB in the natural environment are mainly composed of rare taxa. (A) Proportions of various taxa of HPB in different habitats. Rare taxa (species with a relative sequence abundance $< 0.001\%$ across all samples of a habitat) are shown in red, medium taxa (species with a relative sequence abundance between 0.001% and 0.1% across all samples of a habitat) are shown in blue, and abundant taxa (species with a relative sequence abundance $\geq 0.1\%$ across all samples of a habitat) are shown in purple. (B and C) Relationships between the richness (B) and abundance (C) of HPB and the proportions of rare taxa. Each dot represents a habitat type. In all the depicted scatter plots, the lines indicate the best linear fit, and the shaded areas represent the 95% confidence intervals of the fitted curves. Pearson's correlation tests were used to examine the correlation between the richness and abundance of HPB and the proportion of rare taxa, with P values indicating statistical significance. The richness and abundance values are log-transformed (base 10). (D) Relationship between sequencing depth and the HPB detection rate.

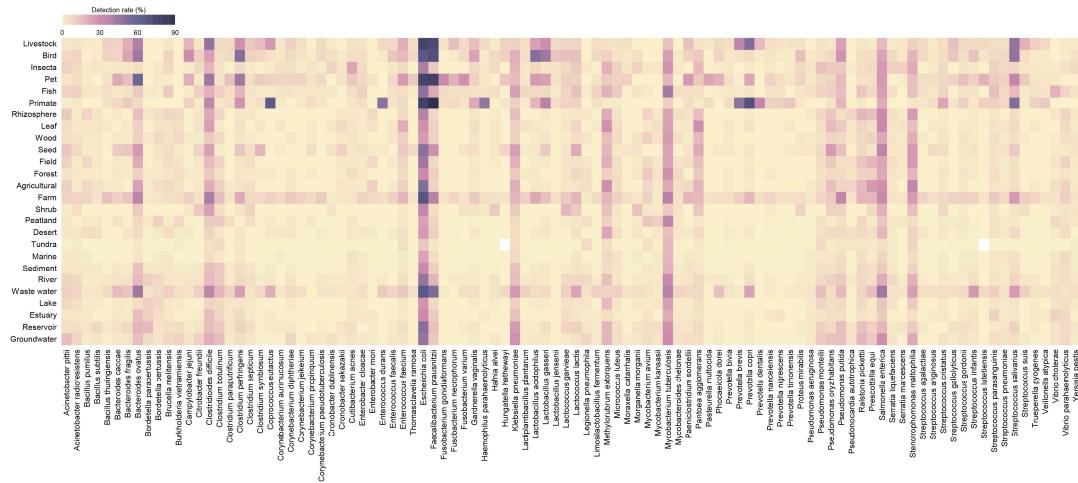


Fig. S5. Detection rates of HPB species in each habitat. The HPB species were selected based on the top 100 average detection rates of 26 habitats.

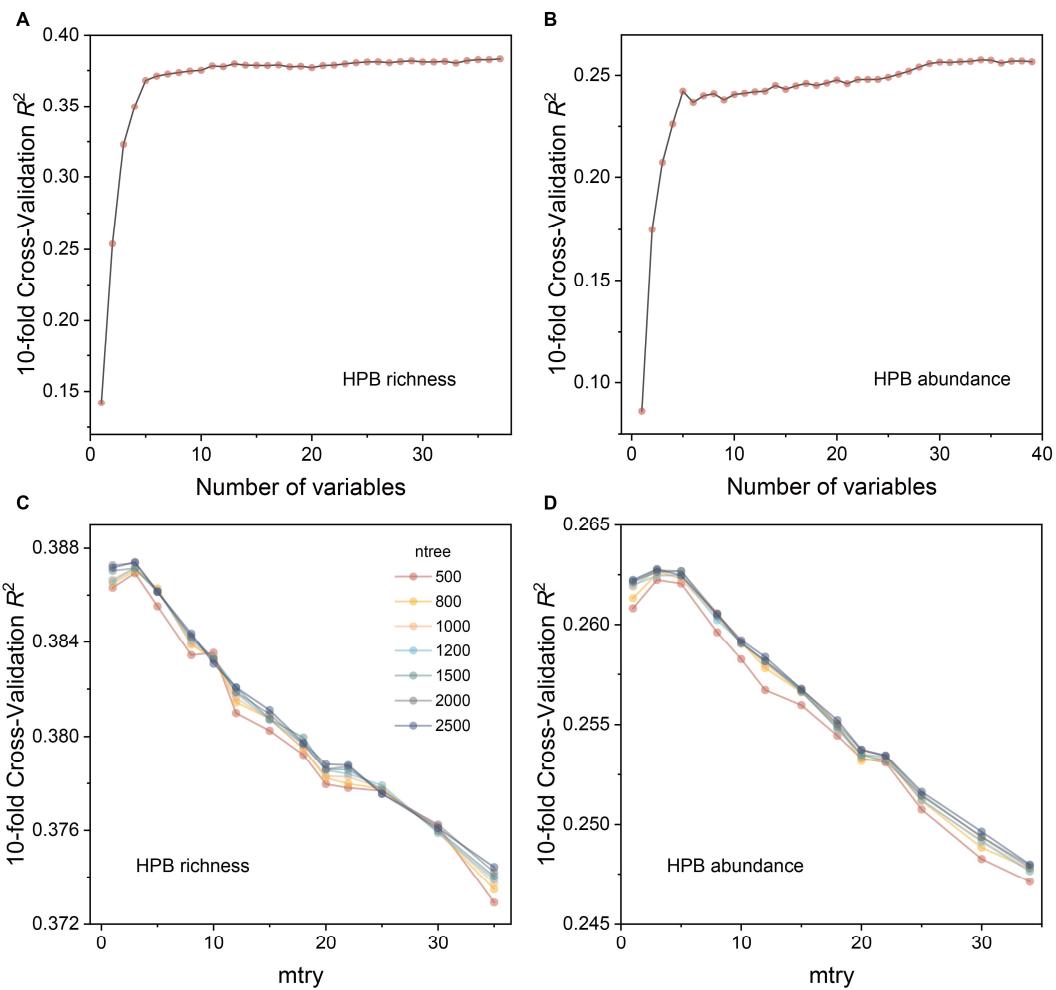


Fig. S6. Feature selection and hyperparameter tuning for machine learning. (A and B) Feature selection for the random forest algorithm was performed using recursive feature elimination and 10-fold cross-validation to identify the optimal set of variables for predicting the richness (A) and abundance (B) of HPB. (C and D) Hyperparameter tuning for the random forest algorithm was conducted using grid search and 10-fold cross-validation to select the optimal combination of hyperparameters for predicting the richness (C) and abundance (D) of HPB.

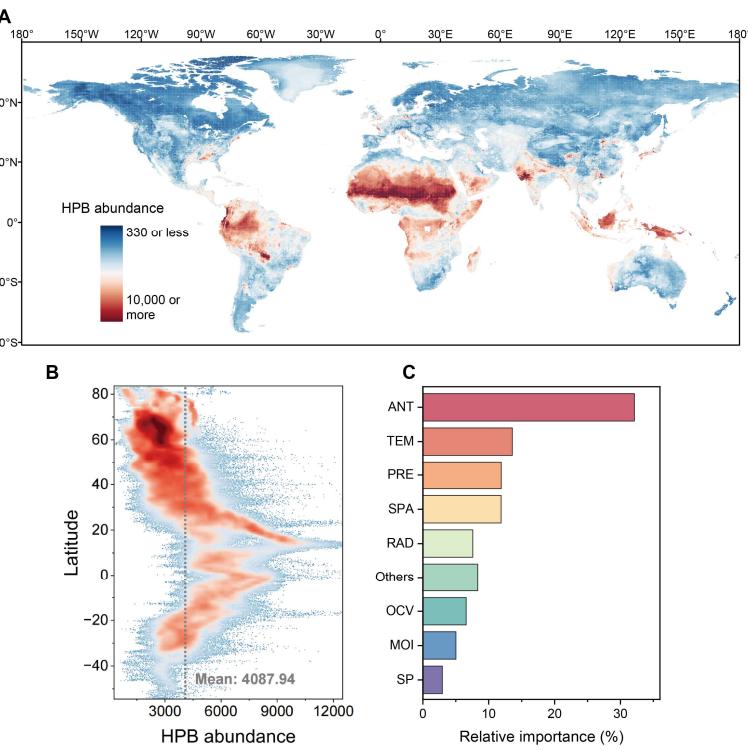


Fig. S7. Global pattern of HPB abundance. (A) Global map of the abundance of HPB. Using covariates, we predicted the abundance of HPB globally based on the random forest model. (B) Latitudinal distribution of the global abundance of HPB. The dashed line represents the average abundance of HPB worldwide. (C) Relative importance of each major category variable in predicting the abundance of HPB. ANT: Anthropogenic, TEM: Temperature, PRE: Precipitation, SPA: Spatial, RAD: Radiation, OCV: Other climatic variables, MOI: Moisture, SP: Soil properties.

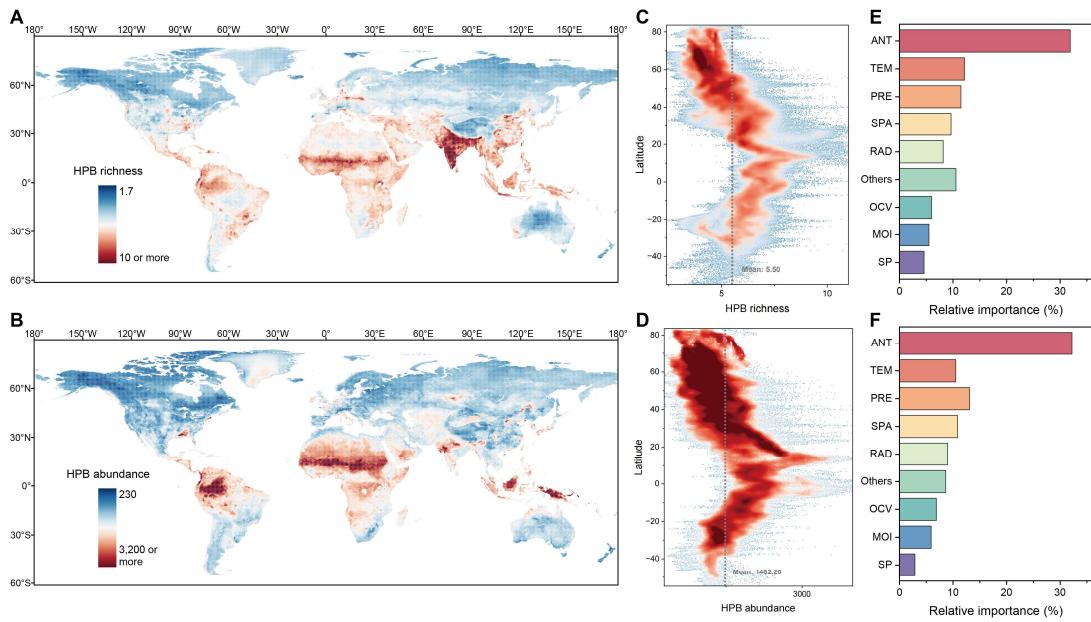


Fig. S8. Global patterns of the richness and abundance of HPB excluding common commensal and opportunistic pathogenic bacteria. (A and B) Global map of the richness (A) and abundance (B) of HPB excluding common commensal and opportunistic pathogenic bacteria. (C and D) Latitudinal distributions of the global richness (C) and abundance (D) of HPB excluding common commensal and opportunistic pathogenic bacteria. The dashed line represents the average richness or abundance of HPB worldwide. (E and F) Relative importance of each major category variable in predicting the richness (E) and abundance (F) of HPB excluding common commensal and opportunistic pathogenic bacteria. ANT: Anthropogenic, TEM: Temperature, PRE: Precipitation, SPA: Spatial, RAD: Radiation, OCV: Other climatic variables, MOI: Moisture, SP: Soil properties. The removed common commensal and opportunistic pathogenic bacteria include: *Escherichia coli*, *Salmonella enterica*, *Faecalibacterium prausnitzii*, *Clostridioides difficile*, *Staphylococcus aureus*, *Enterococcus faecium*, *Prevotella copri*, *Corynebacterium jeikeium*, *Serratia marcescens*, *Legionella pneumophila*, *Mycobacterium avium*, *M. tuberculosis*, *Pseudomonas aeruginosa*, *P. putida*, *Streptococcus pyogenes*, *S. pneumoniae*, and *S. parasanguinis*.

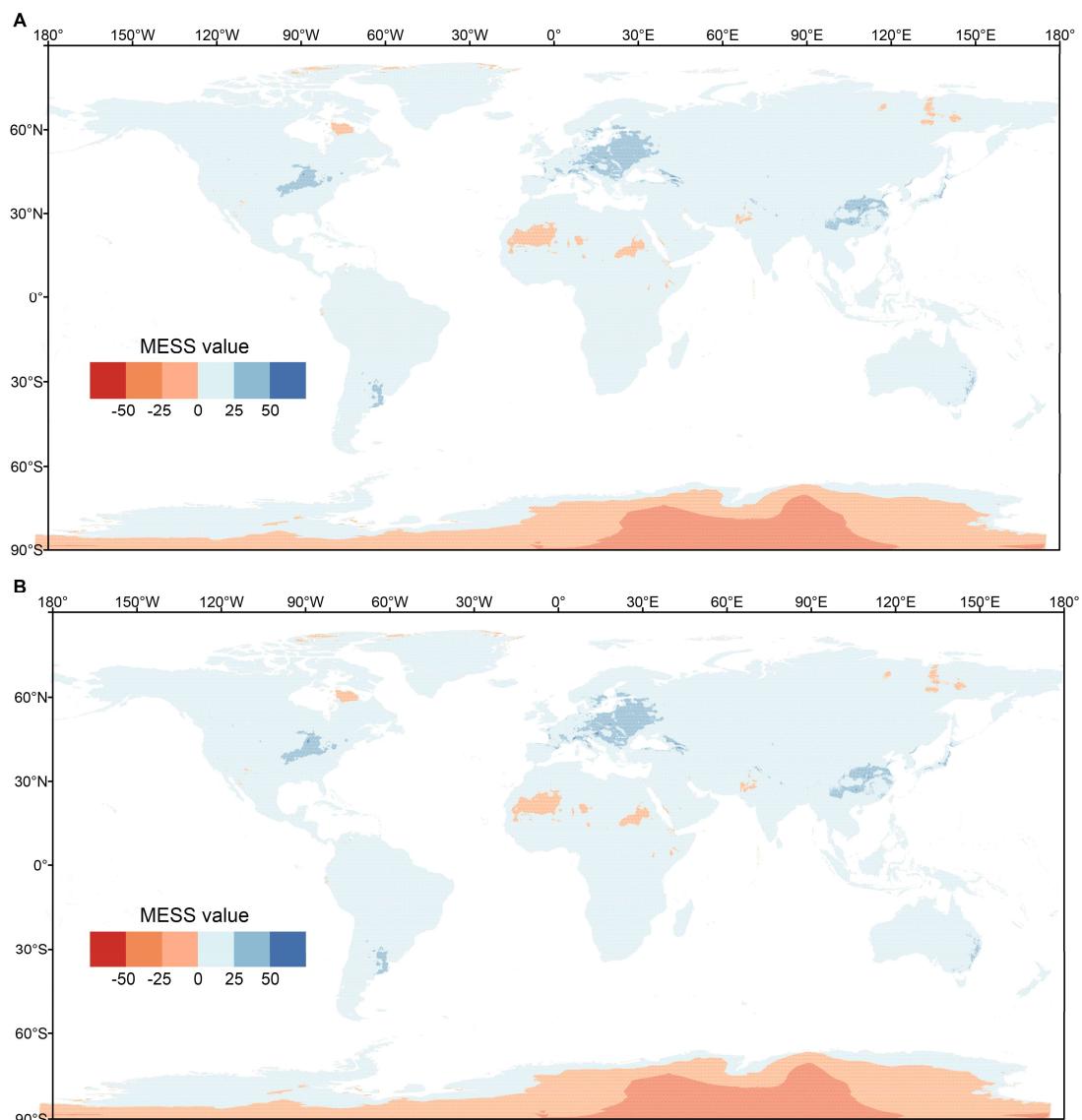


Fig. S9. Multivariate environmental similarity surface (MESS) analysis. (A and B)
Extrapolation reliability of the samples was evaluated through MESS analysis. The higher MESS value indicates high extrapolation reliability.

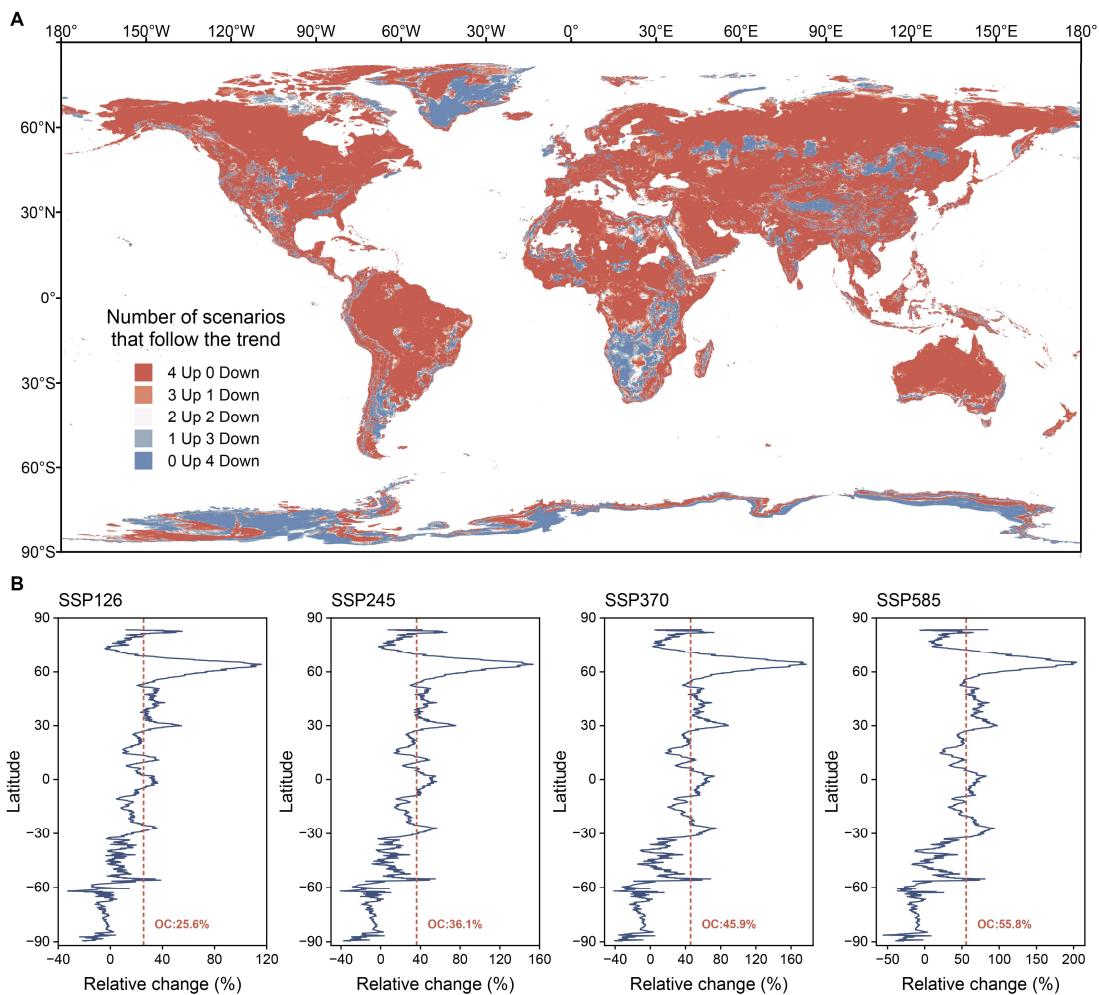


Fig. S10. Abundance of potential HPB under future climate change scenarios. (A) Relative changes in HPB abundance under future climate change scenarios. Based on the historical data of 19 bioclimatic variables, a model was constructed using the random forest algorithm to predict the abundance of HPB under current climate conditions. Using the constructed model, based on future (2080-2100) data of 19 bioclimatic variables, we predicted future HPB abundance under four future climate change scenarios. "Up" represents the number of scenarios in which the HPB abundance increases, whereas "Down" represents the number of scenarios in which the HPB abundance decreases under future climate change scenarios. (B) Latitudinal changes in the abundance of HPB under future climate change scenarios. The dashed line represents the overall change (OC) in HPB abundance under future climate scenarios compared to the current. Shared socioeconomic pathway (SSP) 126, sustainability; SSP245, middle of the road; SSP370, regional rivalry; SSP585, fossil-fuelled development.

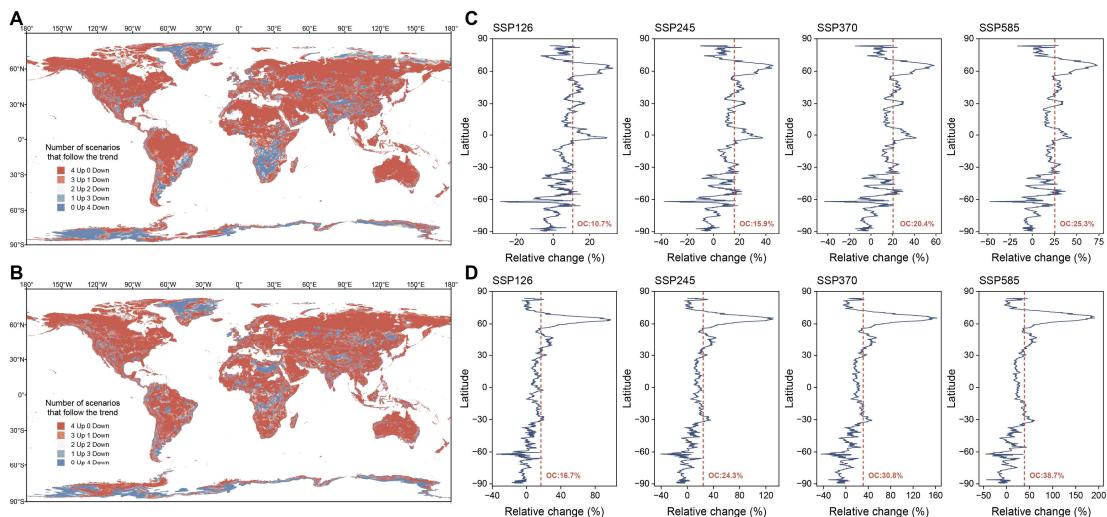


Fig. S11. Richness and abundance of potential HPB excluding common commensal and opportunistic pathogenic bacteria under future climate change scenarios. (A and B) Relative changes in the richness (A) and abundance (B) of HPB excluding common commensal and opportunistic pathogenic bacteria under future climate change scenarios. Based on the historical data of 19 bioclimatic variables, models were constructed using the random forest algorithm to predict the richness and abundance of HPB excluding common commensal and opportunistic pathogenic bacteria under current climate conditions. Using the constructed models, based on future (2080-2100) data of 19 bioclimatic variables, we predicted future the richness and abundance of HPB excluding common commensal and opportunistic pathogenic bacteria under four future climate change scenarios. "Up" represents the number of scenarios in which HPB richness or abundance increases, whereas "Down" represents the number of scenarios in which HPB richness or abundance decreases under future climate change scenarios. (C and D) Latitudinal changes in the richness (C) and abundance (D) of HPB excluding common commensal and opportunistic pathogenic bacteria under future climate change scenarios. The dashed line represents the overall change (OC) in HPB richness or abundance under future climate scenarios compared to the current. Shared socioeconomic pathway (SSP) 126, sustainability; SSP245, middle of the road; SSP370, regional rivalry; SSP585, fossil-fuelled development.

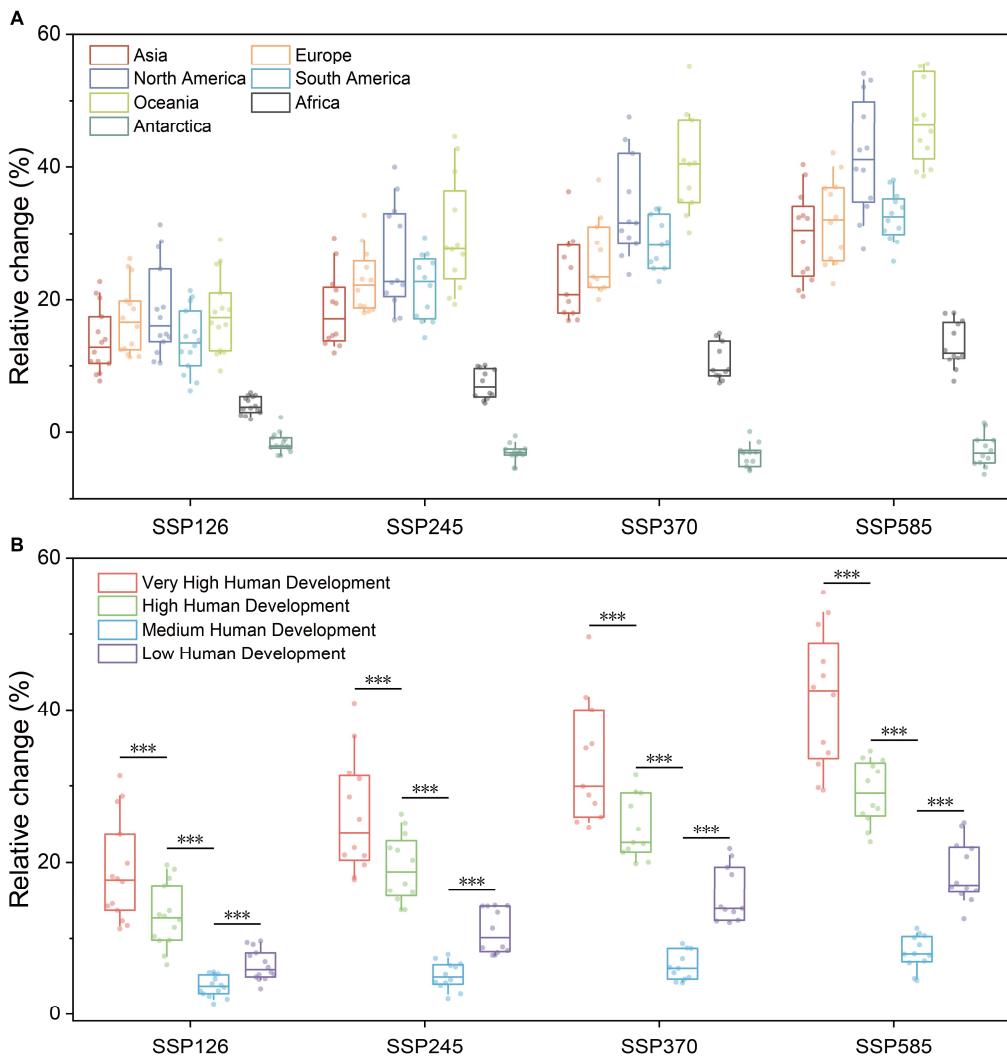


Fig. S12. Relative changes in HPB richness in different regions under future climate change scenarios. (A) Relative changes in the richness of HPB on different continents under future climate change scenarios. (B) Relative changes in the richness of HPB in countries with different human development levels under future climate change scenarios. Comparisons between bins were conducted using the Wilcoxon signed-rank test, *** $P < 0.001$. In all the depicted boxplots, the middle line indicates the median, the box represents the 25th-75th percentiles, and the error bar indicates the 10th-90th percentiles of the observations. Dots represent the changes in richness predicted by different GCMs compared to the current richness.

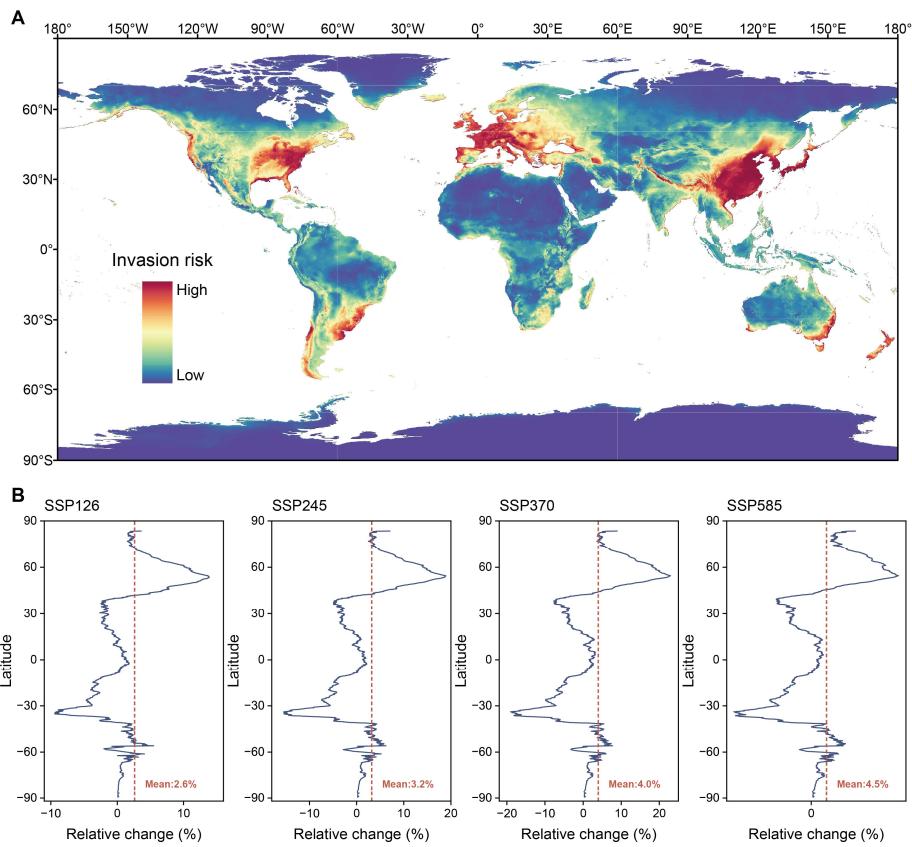


Fig. S13. Global invasion risk of potential HPB. (A) Predicted global invasion risk of potential HPB under current climate conditions. (B) Latitudinal changes in the invasion risk of HPB under future climate change scenarios. The dashed line represents the average change in the invasion risk of HPB under future climate scenarios compared to the current.

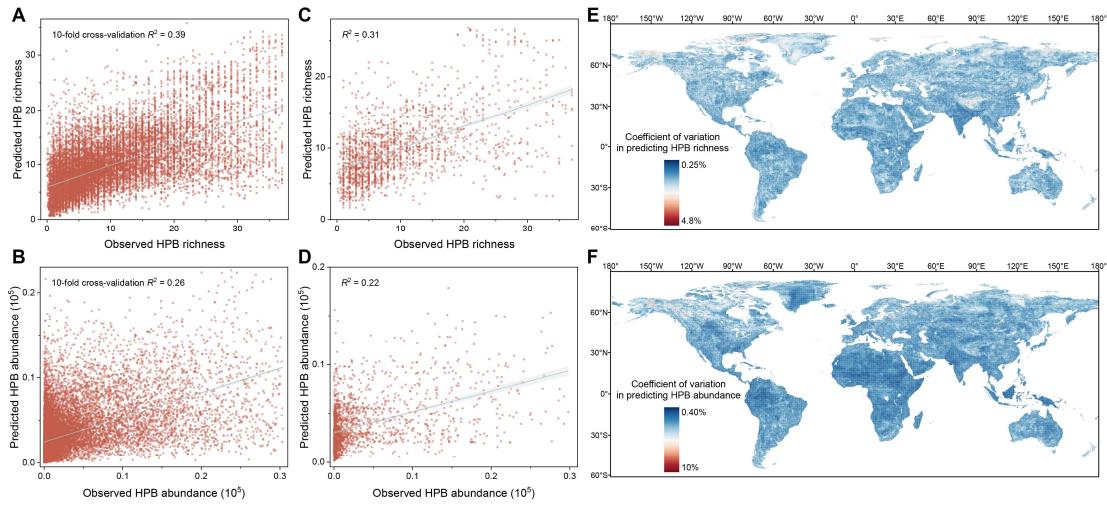


Fig. S14. Evaluation of machine learning models for global distribution. **(A)** Relationship between the predicted and observed richness in the training set. **(B)** Relationship between the predicted and observed abundance in the training set. The performance of the model on the training set was evaluated through 10-fold cross-validation R^2 . **(C)** Relationship between the predicted and observed richness in the testing set. **(D)** Relationship between the predicted and observed abundance in the testing set. The performance of the model on the testing set was evaluated through R^2 . **(E) and (F)** Coefficients of variation were used to evaluate the uncertainty of the predicted global HPB richness (E) and abundance (F).

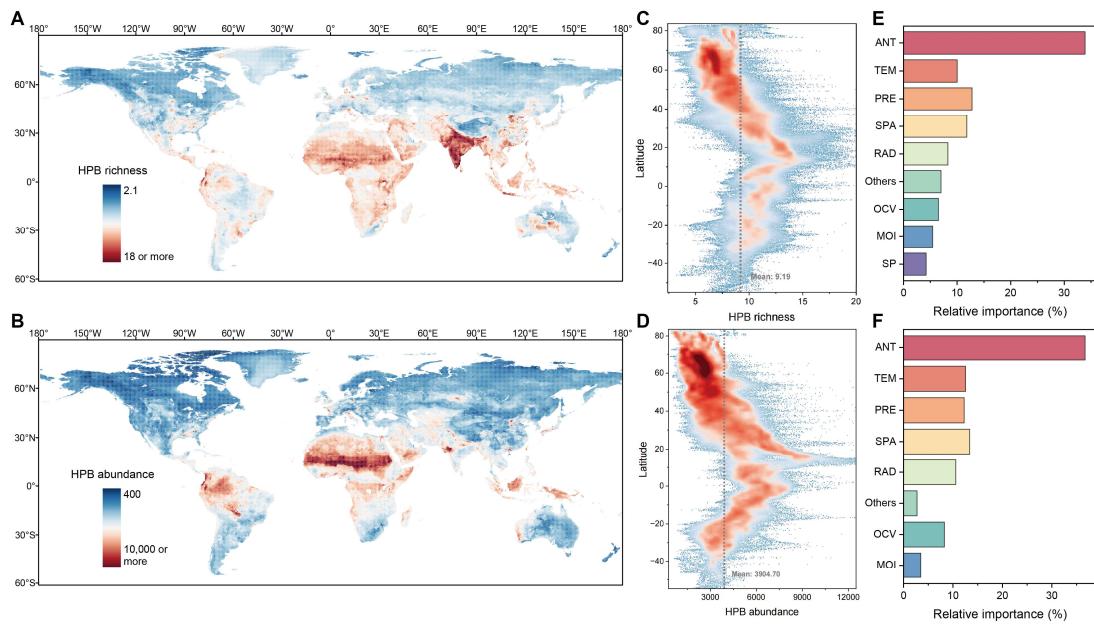


Fig. S15. Global patterns of HPB richness and abundance predicted through spatial cross-validation. (A and B) Global map of the richness (A) and abundance (B) of HPB. Using the random forest algorithm, we predicted the richness and abundance of HPB globally based on spatial cross-validation. (C and D) Latitudinal distributions of the global richness (C) and abundance (D) of HPB. The dashed line represents the average richness or abundance of HPB worldwide. (E and F) Relative importance of each major category variable in predicting the richness (E) and abundance (F) of HPB. ANT: Anthropogenic, TEM: Temperature, PRE: Precipitation, SPA: Spatial, RAD: Radiation, OCV: Other climatic variables, MOI: Moisture, SP: Soil properties.

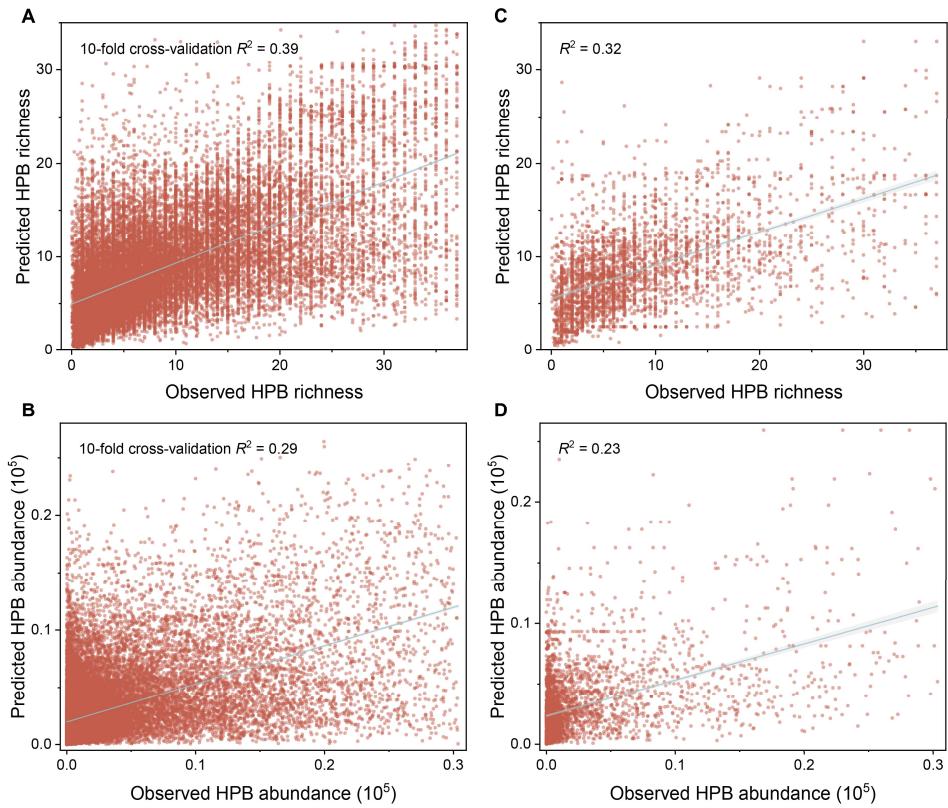


Fig. S16. Evaluation of machine learning models for climate change. **(A)** Relationship between the predicted and observed richness in the training set for climate change. **(B)** Relationship between the predicted and observed abundance in the training set for climate change. The performance of the model on the training set was evaluated through 10-fold cross-validation R^2 . **(C)** Relationship between the predicted and observed richness in the testing set for climate change. **(D)** Relationship between the predicted and observed abundance in the testing set for climate change. The performance of the model on the testing set was evaluated through R^2 .

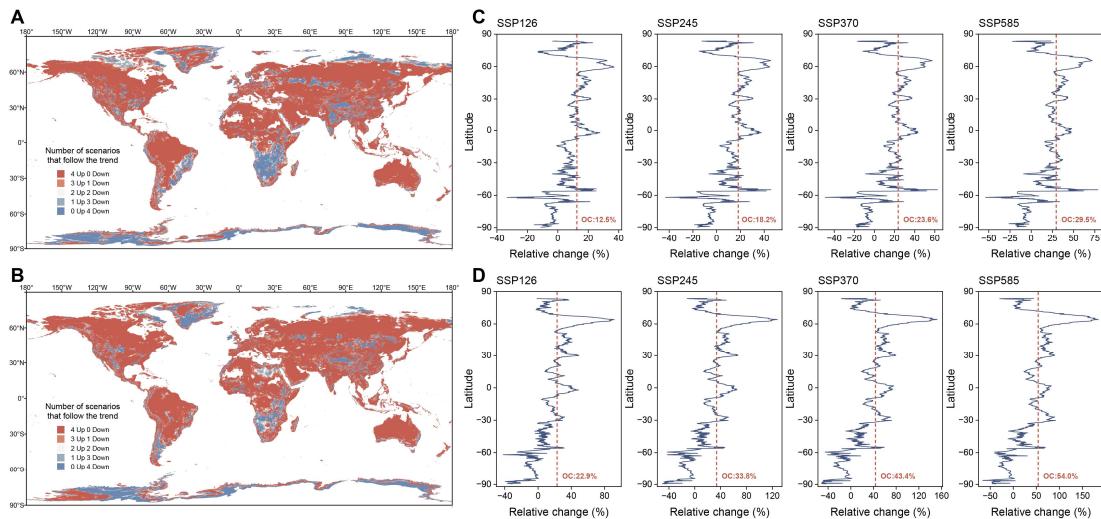


Fig. S17. Richness and abundance of potential HPB predicted through spatial cross-validation under future climate change scenarios. (A and B) Relative changes in HPB richness (A) and abundance (B) under future climate change scenarios. Based on the historical data of 19 bioclimatic variables, models were constructed using the random forest algorithm to predict the richness and abundance of HPB under current climate conditions. The construction of the models is based on spatial cross-validation. Using the constructed model, based on future (2080-2100) data of 19 bioclimatic variables, we predicted future HPB richness and abundance under four future climate change scenarios. "Up" represents the number of scenarios in which HPB richness or abundance increases, whereas "Down" represents the number of scenarios in which HPB richness or abundance decreases under future climate change scenarios. (C and D) Latitudinal changes in the richness (C) and abundance (D) of HPB under future climate change scenarios. The dashed line represents the overall change (OC) in HPB richness or abundance under future climate scenarios compared to the current. Shared socioeconomic pathway (SSP) 126, sustainability; SSP245, middle of the road; SSP370, regional rivalry; SSP585, fossil-fuelled development.

Table S6 Future climate models used in the current study

Model name	SSP126	SSP245	SSP370	SSP585
ACCESS-CM2	✓	✓	✓	✓
BCC-CSM2-MR	✓	✗	✗	✗
CMCC-ESM2	✓	✓	✓	✓
EC-Earth3-Veg	✓	✓	✓	✓
FIO-ESM-2-0	✓	✓	✗	✓
GFDL-ESM4	✓	✗	✓	✗
GISS-E2-1-G	✓	✓	✓	✓
HadGEM3-GC31-LL	✓	✓	✗	✓
INM-CM5-0	✓	✓	✓	✓
IPSL-CM6A-LR	✓	✓	✓	✓
MIROC6	✓	✓	✓	✓
MPI-ESM1-2-HR	✓	✓	✓	✓
MRI-ESM2-0	✓	✓	✓	✓
UKESM1-0-LL	✓	✓	✓	✓

Other Supplementary Materials for this manuscript include the following:

Table S1 Taxonomic information of human pathogenic bacteria

Table S2 Socioeconomic factors of each country

Table S3 The covariates used in machine learning

Table S4 The relative importance of variables in predicting the richness of human pathogenic bacteria

Table S5 The relative importance of variables in predicting the abundance of human pathogenic bacteria