Environmental Science & Technology Letters

pubs.acs.org/journal/estlcu

Letter

Integrating Functional Genomic Descriptors into Explainable Machine-Learning Models for Cross-Aquatic Species Prediction of Metal-Containing Nanomaterials Ecotoxicity

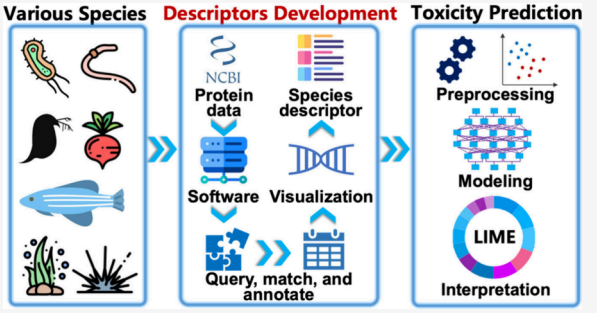
Yunchi Zhou, Ying Wang,* Willie J. G. M. Peijnenburg, Martina G. Vijver, Kenneth M. Y. Leung, Surendra Balraadjsing, Xiaoli Zhao, Wenhong Fan,* and Fengchang Wu

Cite This: *Environ. Sci. Technol. Lett.* 2026, 13, 1114−1121

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Understanding and predicting the ecotoxicity of nanomaterials (NMs) in untested aquatic species remains a major challenge due to the absence of quantitative and mechanistically meaningful species descriptors. Here, we propose a functional genomics-based species descriptor that captures molecular functions, antioxidant defenses, and detoxification capacities and evaluate its utility in improving deep-learning (DL) models for NM toxicity prediction. Using toxicity data for 14 metal-containing NMs across 31 aquatic organisms, we integrated the proposed descriptor into four supervised machine-learning algorithms, the models achieved a highly balanced and satisfactory predictive performance across different structures, with cross-validation Q^2 values stabilizing around 0.61−0.65 for both random forest (RF) and DL architectures. Feature-importance analysis further showed that functional genomic traits contributed comparably to NM physicochemical properties and exposure conditions, indicating that the descriptor provides mechanistic relevance rather than functioning as a purely statistical feature set. This approach enables the quantitative characterization of species-specific sensitivity to NMs and substantially enhances cross-species generalizability of toxicity prediction. Consequently, the proposed genomics-based descriptor provides a scalable approach for hazard assessment of NMs in untested, rare, and data-deficient aquatic species of conservation concern.



KEYWORDS: *nanomaterials, toxicity, cross-species, functional genomics, species descriptors, deep learning, prediction*

INTRODUCTION

The rapid development and diversification of engineered nanomaterials (NMs) pose major challenges for ecological risk assessment.[1,2] Conventional *in vivo* and *in vitro* toxicity testing is resource-intensive and often insufficient to keep pace with the growing number of newly produced NMs.[3] Consequently, *in silico* approaches, particularly quantitative structure-activity relationship (QSAR) modeling and machine learning (ML), have gained traction as complementary tools capable of accelerating nanotoxicity prediction while supporting hypothesis-driven experimental design.[4−6] Early nano-QSAR efforts primarily relied on NM physicochemical descriptors, and subsequent advances have incorporated quantum chemical features, periodic table parameters, and mechanistically grounded descriptors such as metal-ligand binding characteristics.[7] These developments have expanded the predictive capacity of computational nanotoxicology and enabled models to handle increasingly complex and nonlinear relationships.

Despite these methodological advances, a fundamental limitation persists: most existing models remain *material-centric*, encoding detailed information on NM properties and exposure conditions while treating biological variability in an overly simplified or indirect manner.[8−10] In cross-species applications, biological differences are commonly represented using taxonomic categories, phylogenetic distance, or coarse ecological and life-history traits within read-across or multi-species modeling frameworks.[11,12] While these methods provide a pragmatic means of extrapolating toxicity thresholds, they implicitly assume that taxonomic relatedness or broad phenotypic similarity adequately reflects shared toxicity mechanisms. Such assumptions are often violated for NMs, whose modes of action are highly context-dependent and mediated by molecular-level interactions. As a result, taxonomy- or trait-based models frequently lack mechanistic

Received: June 15, 2026
Revised: July 15, 2026
Accepted: July 16, 2026
Published: July 25, 2026



Downloaded from pubs.acs.org/estlcu/article-pdf/13/8/1114/6625613/acs.estlett.6c00579.pdf by BEIHANG UNIV user on 18 August 2026

ACS Publications

© 2026 American Chemical Society

https://doi.org/10.1021/acs.estlett.6c00579
Environ. Sci. Technol. Lett. 2026, 13, 1114−1121

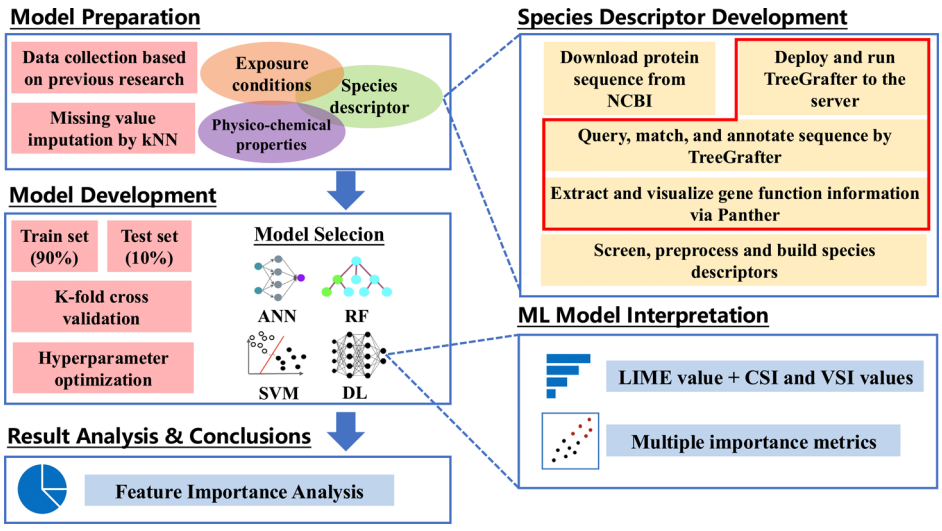


Figure 1. Workflow of the development of species descriptors. The content in the red solid box is completed by self-developed software tools: *k*NN, *k* nearest neighbor; ANN, artificial neural network; RF, random forest; SVM, support vector machine; DL, deep learning; LIME, Local Interpretable Model-agnostic Explanations; VSI, variable stability index; CSI, coefficient stability index.

resolution and struggle to generalize across distantly related taxa or novel species.

Our previous review found that fewer than one-third of ML-based nanotoxicology studies incorporated species-level information,[13] and these predominantly focused on *in vitro* systems such as cell types or cell lines.[13−17] Only a small subset of studies attempted to address interspecies sensitivity using *in vivo* endpoints. Efforts to integrate species traits—such as trophic level or phylogenetic distance—have demonstrated potential for improving cross-species predictions, yet they also highlight the challenge of obtaining quantitative, mechanistically meaningful descriptors that can generalize across diverse taxa.[18,19]

The need for biologically informed species descriptors is particularly relevant to nanotoxicology. The toxicity of NMs arises from interactions among their physicochemical characteristics, environmental conditions, and organism-specific biological responses. These organismal responses, such as antioxidant defense, detoxification, and molecular stress pathways, are rooted in the genomic function. The increasing availability of genome-wide data across aquatic taxa now provides an opportunity to incorporate functional biological information into predictive models.[20] Functional genomics offers mechanistic insights into gene products, pathways, and stress-response functions that fundamentally shape species sensitivity to contaminants.[21] Although functional genomics has been extensively used in biomedical research to elucidate molecular mechanisms of toxicity, its potential to serve as a structured and computable species descriptor in QSAR- or ML-based nanotoxicity prediction modeling remains unexplored.[22]

Here, we address this knowledge gap by proposing that appropriately processed functional genomic information can serve as an interpretable, mechanistically grounded, and machine-readable species descriptor. Such a descriptor could (i) capture interspecies variation in biological processes relevant to NM toxicity, (ii) improve the generalizability of predictive models across species, and (iii) enable hazard estimation for untested, rare, or data-deficient organisms. To our knowledge, no previous studies have integrated species functional genomics into computational nanotoxicology. This work therefore aims to develop a new genomics-based species descriptor and incorporate it into ML models of NM toxicity. We outline an approach to obtain, annotate, and transform genomic information into quantitative descriptors, integrate these descriptors with established nanotoxicity features, and evaluate their utility within optimized modeling workflows. Ultimately, this study seeks to determine whether functional genomics can meaningfully enhance the predictive capacity and mechanistic interpretability of nanotoxicity models and to provide a novel approach enabling read-across to untested aquatic species, including rare and endangered species.

MATERIALS AND METHODS

2.1. Collection and Processing of Genomic Information

Gene ontology (GO) provides a standardized framework for functional classification of homologous genes across species, including three domains: molecular function (MF), biological process (BP), and cellular component (CC).[23] Because comprehensive functional genomic data are unavailable for most aquatic organisms, we constructed species-level functional genomics data independently.

Protein sequence files (.faa) for each species were retrieved from the National Center for Biotechnology Information (NCBI).[24] These sequences were annotated using TreeGrafter, a phylogeny-aware functional annotation approach that assigns proteins to pre-annotated gene family trees and infers functional annotations through ancestral state reconstruction. Compared with direct homology-based methods (e.g., sequence similarity searches), phylogenetic placement has been shown to improve functional inference accuracy by reducing annotation transfer errors arising from paralogy and sequence divergence, particularly for non-model organisms. A detailed justification for the selection of TreeGrafter, including comparison with alternative annotation strategies and supporting literature on annotation accuracy, is provided in Text S1.[25] To overcome TreeGrafter's lack of multispecies data management and visualization features, we developed an integrated pipeline enabling batch processing, annotation, and visualization of functional genomic profiles (Text S1 and Figure 1).

Annotated genes were then classified using PANTHER into GO categories and protein class (PC). Functional gene counts under each category were extracted and normalized by the total gene number to produce species-level functional genomic profiles. Gene functions absent across all species or present at negligible levels (≤0.1%) were excluded (Text S2).

https://doi.org/10.1021/acs.estlett.6c00579
Environ. Sci. Technol. Lett. 2026, 13, 1114−1121

Downloaded from pubs.acs.org/estlcu/article-pdf/13/8/1114/6625613/acs.estlett.6c00579.pdf by BEIHANG UNIV user on 18 August 2026

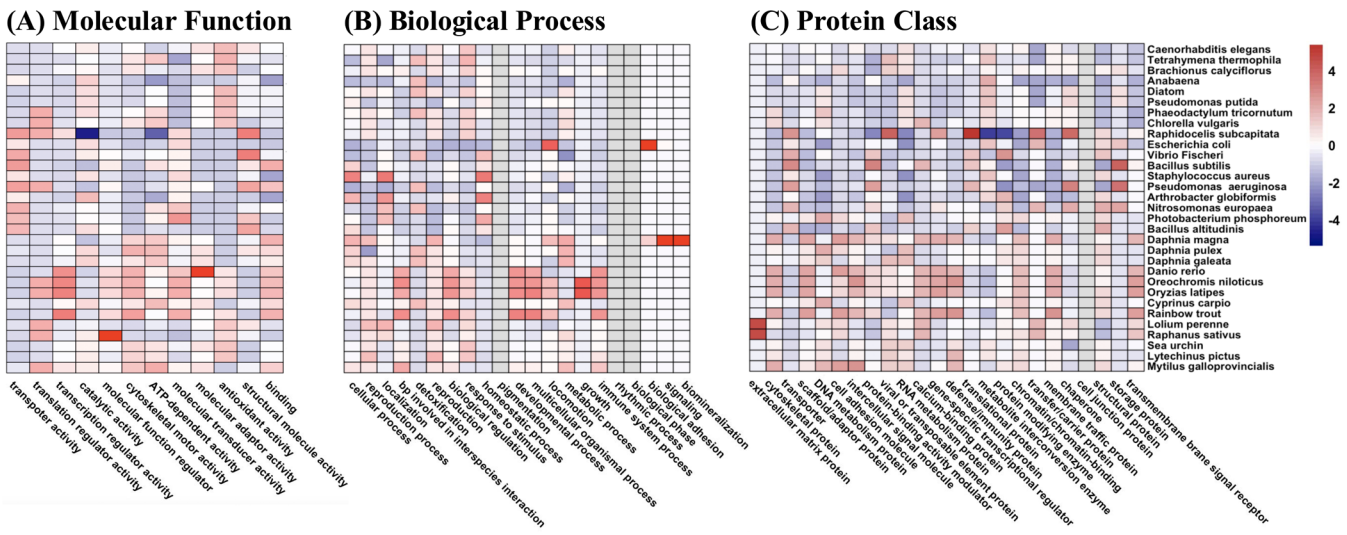


Figure 2. Heatmap of the multispecies descriptor, illustrating the variations in functional gene abundance across different species: (A) MF; (B) BP; (C) PC. Each block represents a classification method, the horizontal bars represent different gene functions, and the vertical bars represent different species. The color intensity reflects the relative proportion of each functional gene after normalization; higher proportions are indicated by more intense red hues, while gray blocks signify that the corresponding gene function is absent in that species.

2.2. Development of Species Descriptors and Data Preprocessing

From the five PANTHER classification systems, MF, BP, and PC were selected for constructing species descriptors because they capture biological processes most closely associated with toxic responses (e.g., oxidative stress and detoxification). For each species, the proportion of genes in each functional category was used as the descriptor. Missing functional categories were assigned zero. Toxicity data were obtained from Nano E-Tox[26] and published databases,[27,28] with EC_{50} (mg/L) used as the acute toxicity endpoint. Species lacking genomic information or NCBI protein sequences were excluded. All numerical descriptors, including functional genomics-based descriptors, nanospecific physicochemical properties, exposure condition parameters, and molecular descriptors, were standardized prior to modeling. To further mitigate the risk of high dimensionality, sequential backward selection (SBS) was implemented to refine the genome descriptor (Text S3). Detailed definitions of the last three descriptor categories can be found in Table S1. Missing values were imputed using k nearest neighbors (kNNs), which preserves the local structure and intervariable relationships more effectively than mean imputation.

2.3. Model Development, Validation, and Importance Analysis

Four supervised learning algorithms, i.e., random forest (RF), support vector machine (SVM), artificial neural network (ANN), and deep learning (DL), were developed following OECD principles for modeling. Hyperparameters were optimized manually to ensure an optimal model performance, with the RF model employing 600 trees and seven features per split, the SVM model using a radial basis function kernel, the ANN model training with the Adam optimizer at a learning rate of 0.1 with 38 hidden units, and the DL model constructing with three hidden layers (256−256−128) and training by using the Adam optimizer with mean-squared error (MSE) as the loss function (detailed in Text S4). The model performance was evaluated using the coefficient of determination (R^2) and the root-mean-square error (RMSE). Python 4.0 with TensorFlow, scikit-learn, and Keras were used for model development.

A 10-fold cross-validation test was used for internal validation, and the resulting Q^2 was used to assess the predictive accuracy of the model. A 90/10 train-test split was used for external validation, with R^2_{ext} and $RMSE_{ext}$ as evaluation metrics. All validation workflows and additional performance comparisons are provided in Text S5.

Feature importance was evaluated by combining tree-based importance metrics, including the minimum depth, P value, node purity increase, and MSE increase, with model-agnostic interpretability methods.[18] For nontree DL models, repeated local interpretable model-agnostic explanations (LIME) interpretation methods were performed, and the stability was quantified using the variable stability index (VSI) and coefficient stability index (CSI). LIME operates on the principle of approximating the behavior of complex models within a specific prediction range by training a simpler, interpretable surrogate model, thereby helping to explain why the complex model makes certain predictions.[29] A composite feature ranking was generated by averaging stable LIME outputs. The detailed LIME procedures and stability analyses are provided in Text S6.

2.4. Application to Untested Aquatic Species

To evaluate the model's cross-species applicability beyond the training taxa, an independent external validation data set comprising five diverse aquatic species—including fish (*Oreochromis niloticus* and *Oryzias melastigma*), crustaceans (*Daphnia similis*), and various microalgae/cyanobacteria (*Thalassiosira pseudonana* and *Synechocystis* sp.)—was compiled. The validation set utilized median effective concentrations (EC_{50}) paired with high-quality genome and proteome annotations retrieved from the NCBI for the exact same species. Detailed metadata for the external validation set and the corresponding performance metrics are provided in Text S7.

RESULTS AND DISCUSSION

3.1. Functional Genomics-Based Species Descriptors

We compiled functional genomic annotations for 31 aquatic species represented in the NMs toxicity data set. To mitigate the risk of high dimensionality, we combined the SBS algorithm with expert toxicological screening to refine the initial 50 descriptors into 25 high-relevance genomic features (Table S2). A detailed distribution report across the species can be found in Table S3. To mitigate species-specific bias, we replaced categorical identities with functional genomic descriptors. This approach shifts the model's focus from individual species traits to conserved biological pathways. By leveraging the evolutionary conservation of these functions, the model can generalize patterns from data-rich species to a

Downloaded from pubs.acs.org/estlcu/article-pdf/13/8/1114/6625613/acs.estlett.6c00579.pdf by BEIHANG UNIV user on 18 August 2026

https://doi.org/10.1021/acs.estlett.6c00579
Environ. Sci. Technol. Lett. 2026, 13, 1114−1121

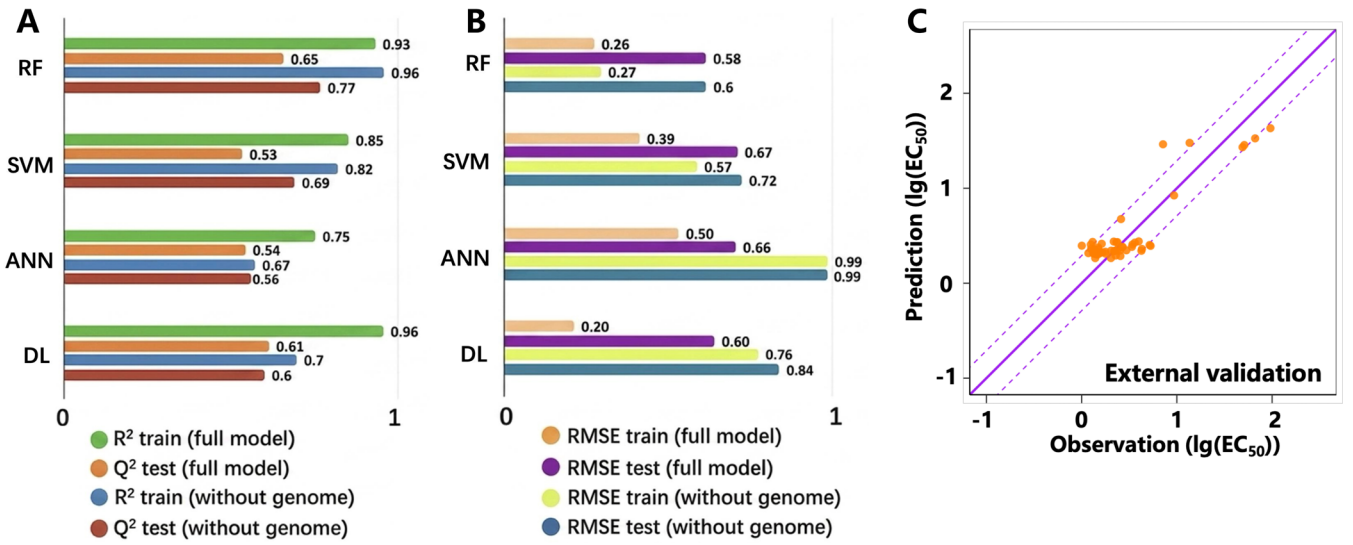


Figure 3. Comparison of the ML model performance with and without the inclusion of species descriptors and validation results of untested species. (A) Comparison of R^2 and Q^2 of RF, SVM, ANN, and DL with and without species descriptors. (B) Comparison of the RMSE of RF, SVM, ANN, and DL with and without species descriptors. (C) Validation result of the RF model on untested aquatic species. The slope of the solid line is 1, and the dotted lines represent the intercepts ± RMSE. R^2: coefficient of determination in the training set. Q^2: coefficient of determination in the test set. RMSE: root-mean-square error.

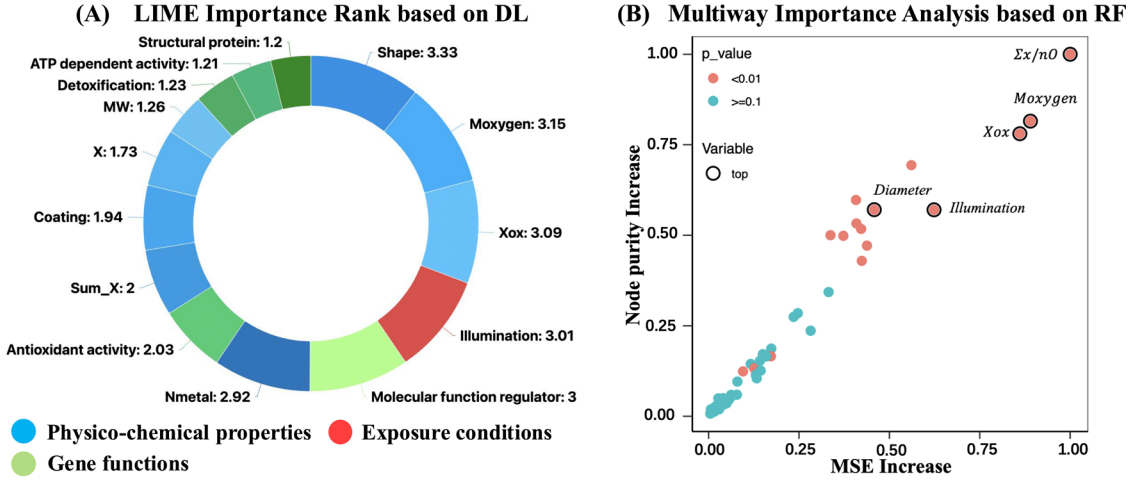


Figure 4. Feature importance analysis of models. (A) LIME importance rank of different descriptors. The features are arranged clockwise in descending order of importance, with colors representing their respective functional categories. (B) Multiway feature importance analysis of models, combining the MSE increase, node purity increase, and *P* value of the features. The *x* axis represents the MSE increase, where higher values indicate greater feature importance; the *y* axis represents the increase in node purity, with larger values signifying a higher contribution to model accuracy. The color scale denotes the *P* value, where features with *P* values of <0.01 are highlighted in red, indicating that the observed importance is statistically significant and highly unlikely to occur under the null hypothesis.

broader range of aquatic taxa, ensuring robust cross-species predictions even for underrepresented organisms.

The overview of detailed species genomic functions can be found in Figures S1−S3. Despite the distribution diversity across taxa, species within the same biological group showed coherent functional profiles (Figure 2). This pattern supports the premise that genomic function composition can capture fundamental similarities in stress-response capacities,[30] thereby enabling read across of NM toxicity to species lacking empirical data.

The functional genomic descriptors proposed here represent species-intrinsic characteristics, distinct from dynamic, exposure-induced transcriptomic or proteomic responses. This stability enables the assessment of NM toxicity across a wide range of untested species without resource-intensive, case-by-case omics experiments, offering significant advantages for high-throughput prediction. Nevertheless, certain limitations warrant acknowledgment. First, variations in the genome assembly and annotation quality across species—particularly between model and nonmodel organisms—may introduce bias; a "0" value in our feature matrix could stem from incomplete sequencing rather than genuine gene absence. To mitigate this, we employed TreeGrafter, a phylogeny-based tool more robust than traditional alignment in identifying orthologues across divergent sequences. Second, our model assumes that gene counts serve as a proxy for an organism's "genetic potential" to cope with stressors, a simplification that overlooks complex regulatory mechanisms such as gene expression, post-translational modifications, and enzyme kinetics. Despite these biological simplifications, this method

https://doi.org/10.1021/acs.estlett.6c00579
Environ. Sci. Technol. Lett. 2026, 13, 1114−1121

Downloaded from pubs.acs.org/estlcu/article-pdf/13/8/1114/6625613/acs.estlett.6c00579.pdf by BEIHANG UNIV user on 18 August 2026

aims to provide a pragmatic and scalable framework for cross-species toxicity prediction, the efficacy of which is further evaluated in the following sections.

3.2. Model Performance and Effects of Species Descriptors

A data set of 628 toxicity records covering 14 metal-containing NMs was assembled, integrating physicochemical properties, molecular descriptors, exposure conditions, and species descriptors. Four ML regression models—RF, SVM, ANN, and DL—were trained with and without species descriptors.

All models achieved acceptable predictive power following the feature reduction (Figure 3). While the RF model exhibited a highly balanced and robust performance profile (R^2 = 0.93 and Q^2 = 0.65), suggesting its reliability as a benchmark, the DL model achieved a resilient predictive capacity under the refined 25-descriptor feature space (R^2 = 0.96 and Q^2 = 0.61). This dual-model verification—leveraging RF for stability and DL for capturing complex nonlinear interactions—ensures that our framework captures biologically meaningful patterns rather than relying on empirical data fitting. The stabilization of Q^2 values within the 0.61−0.65 range across different algorithmic structures demonstrates that the framework possesses powerful and generalizable capabilities for cross-species toxicity prediction, successfully shifting the predictive focus from organism-specific classifications to shared core biological machineries.

The observed versus predicted plots for all four algorithms are presented in Figure S4, further illustrating the comparative predictive behaviors of the models. The tight clustering of predictions within the updated validation limits confirmed that this framework trades redundant empirical correlations for genuine biologically explanatory power, successfully breaking the taxonomic barrier in cross-species toxicity extrapolation. External validations further demonstrated the robustness of the DL model (Figure S5).

These findings indicate that functional genomic information captures biologically meaningful differences among species that are relevant to NM-induced toxicity and provides a measurable improvement beyond physicochemical and molecular descriptors alone.

3.3. Key Drivers of NM Toxicity

Combined feature importance analysis from both the LIME-based DL model (Figure 4A) and multiway feature importance analysis-based RF model (Figure 4B) revealed that NM physicochemical properties (e.g., shape), exposure conditions (e.g., illumination), and molecular descriptors (e.g., oxygen- or metal-related parameters) were consistently among the most influential predictors. These factors are well-known to modulate dissolution behavior, redox activity, and organism−particle interactions, and the present results are therefore consistent with the established mechanistic understanding of NM toxicity.[18,31−33] The influence of shape is linked to differences in the specific surface area and cellular interaction pathways; for instance, rod-shaped NMs may exhibit different membrane-penetration efficiencies or surface reactivity compared to spherical ones, thereby modulating their dissolution and toxic potential.[34] Although dissolution is a well-recognized driver of metallic NM toxicity, its moderate importance ranking in this study likely stems from the high proportion of missing values in the experimental records and the binary representation, which limits the model's ability to fully capture its statistical contribution. Furthermore, while the hydrodynamic diameter remains a fundamental characteristic, its impact in our model appears to be integrated into more complex, oxygen- and metal-related molecular descriptors.[35] These parameters are strongly correlated with the potential for reactive oxygen species generation and the release of toxic metal ions, respectively, providing a clearer mechanistic link between these static descriptors and the observed ecotoxicological effects.[36]

Notably, several species descriptors—including MF regulation, antioxidant activity, and detoxification—were also ranked among the top contributors to model performance. These descriptors represent baseline biological capacities rather than exposure-induced responses, suggesting that the inherent genomic functional composition plays a measurable role in shaping species-specific sensitivity to NMs.[37,38] Partial dependence plots further revealed systematic monotonic relationships between specific functional descriptors and predicted toxicity (Figure S6). In particular, the increased relative abundance of antioxidant and detoxification functions was associated with a higher predicted EC_{50}, suggesting a potential decrease in sensitivity to NMs exposure, as supported by recent studies on the toxic effects of NM.[39,40] This trend is biologically consistent with enhanced capacities for oxidative stress buffering and toxicant processing.[41−44] In contrast, ATP-dependent activity exhibited a strong negative association with predicted EC_{50}, suggesting that species characterized by elevated energy-dependent MFs may be intrinsically more vulnerable to NM-induced perturbations, potentially due to higher metabolic demand or increased susceptibility of energy-coupled pathways.

From a modeling perspective, the emergence of species descriptors among top-ranked features indicates that incorporating functional genomics information reduces the reliance of predictions solely on NM-centric properties and exposure conditions. Instead, the model captures toxicity as an emergent property of NM characteristics interacting with species-specific biological capacity. This shift improves interpretability and supports the use of static genomic descriptors as mechanistically meaningful variables for cross-species toxicity predictions. Extended mechanistic discussion of these relationships within an adverse outcome pathway framework is provided in Text S8.

3.4. Model Validation on Untested Aquatic Species

To further verify the model's applicability on untested species, we performed an external validation on an independent data set comprising five diverse aquatic taxa (*Oreochromis niloticus*, *Oryzias melastigma*, *Daphnia similis*, *Thalassiosira pseudonana*, and *Synechocystis* sp.), which are absent from the training set. The result demonstrated that both the DL and RF models achieved reliable predictive performances; the DL model reached a predictive R^2 of 0.51 and a RMSE of 0.42. Notably, the RF model exhibited superior and more robust generalizability, achieving a R^2 of 0.78 and a RMSE of 0.29 (Figure 3C). This performance signifies that our model can effectively predict the toxicity variance for species that have no prior empirical data. The close alignment between predicted and observed EC_{50} values across different biological groups confirms that functional genomic descriptors can capture universal stress-response signatures. Consequently, this framework represents a significant advancement toward enabling high-throughput ecological risk screening for data-poor, threatened, or conservation revelent aquatic organisms.

Downloaded from pubs.acs.org/estlcu/article-pdf/13/8/1114/6625613/acs.estlett.6c00579.pdf by BEIHANG UNIV user on 18 August 2026

https://doi.org/10.1021/acs.estlett.6c00579
Environ. Sci. Technol. Lett. 2026, 13, 1114−1121

This study demonstrates that incorporating functional genomics-based descriptors significantly enhances the biological interpretability and mechanistic robustness of NM toxicity modeling across species. While absolute statistical outcomes are comparable to our previous empirical findings utilizing traditional trophic levels, this framework represents a fundamental paradigm shift from empirical data fitting to mechanistic inference. By transforming a species identity from a coarse taxonomic placeholder to 25 precise features, the model moves beyond "memorizing" historical group sensitivities. Because these descriptors utilize curated genome annotations rather than resource-intensive omics profiling, they successfully break taxonomic barriers, enabling scalable and biologically explanatory toxicity predictions for rare or untested taxa with baseline genomic data.

Future work should focus on refining genomic descriptors to better understand the functions most relevant to NMs toxicity and expand the experimental data, particularly for underrepresented species and materials. Integrating additional species-level descriptors—such as physiological traits or environmental DNA signatures—may further improve extrapolation to untested taxa. Overall, the proposed approach provides a promising tool toward more inclusive ecological risk assessments for NMs, particularly for data-poor or conservation-relevant species.

■ ASSOCIATED CONTENT

Data Availability Statement

The code for modeling is also available on https://github.com/BUAA-JonasZhou/Genome. The data set for modeling and the species descriptor used in this study are available on our group's Website: https://ainanotox.com/. The code and data set for modeling and validation is publicly available on Zenodo at 10.5281/zenodo.18410892.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.estlett.6c00579.

Detailed methods for genomic data collection, protein sequence annotation, and functional classification (Texts S1−S8); descriptive statistics and construction workflow of functional genomics-based species descriptors (Figures S1−S3); missing value imputation procedures and data quality assessment protocols; ML model hyperparameters, optimization, and selection criteria (Table S1); comprehensive internal (10-fold cross-validation) and external (90/10 split) validation metrics (Figure S4 and Table S2); feature importance analysis combining tree-based metrics and LIME interpretation (Figures S5 and S6); independent external validation metadata and predictions for five untested aquatic species (Table S3) (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Ying Wang − *School of Materials Science and Engineering, Beihang University, Beijing 100191, China;* orcid.org/0000-0002-1424-0422; Email: yingw@buaa.edu.cn, wy2012bnu@126.com

Wenhong Fan − *School of Materials Science and Engineering, Beihang University, Beijing 100191, China;* orcid.org/0000-0001-5047-1714; Email: fanwh@buaa.edu.cn

Authors

Yunchi Zhou − *School of Materials Science and Engineering, Beihang University, Beijing 100191, China;* orcid.org/0000-0001-6603-2177

Willie J. G. M. Peijnenburg − *Institute of Environmental Science, Leiden University, Leiden 2333 CC, The Netherlands; Center for Safety of Products and Substances, National Institute of Public Health and the Environment, Bilthoven 3721 MA, The Netherlands;* orcid.org/0000-0003-2958-9149

Martina G. Vijver − *Institute of Environmental Science, Leiden University, Leiden 2333 CC, The Netherlands*

Kenneth M. Y. Leung − *State Key Laboratory of Marine Pollution and Department of Chemistry, City University of Hong Kong, Hong Kong 999077, China;* orcid.org/0000-0002-2164-4281

Surendra Balraadjsing − *Institute of Environmental Science, Leiden University, Leiden 2333 CC, The Netherlands;* orcid.org/0000-0003-2167-8163

Xiaoli Zhao − *State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing 100012, China;* orcid.org/0000-0002-5671-389X

Fengchang Wu − *State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing 100012, China;* orcid.org/0000-0001-8595-2243

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.estlett.6c00579

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Nos. 42330710 and 42177240), Beijing Natural Science Foundation (No. 8242033), the National Key R&D Program of China (No. 2022YFC3204800), and the Fundamental Research Funds for the Central Universities. S.B. and M.G.V. received funding from the European Union's ERC-consolidator Grant Agreement No. 101002123.

■ REFERENCES

(1) Basei, G.; Hristozov, D.; Lamon, L.; Zabeo, A.; Jeliazkova, N.; Tsiliki, G.; Marcomini, A.; Torsello, A. Making use of available and emerging data to predict the hazards of engineered nanomaterials by means of in silico tools: A critical review. *NanoImpact* **2019**, *13*, 76−99.

(2) Zhang, D.; Ren, B.; Liu, H.; Li, C.; Wang, X.; Fan, W. Challenges in assessing ecological and health risks of microplastics and nanoplastics: tracking their dynamics in living organisms. *New Contaminants* **2026**, *2* (1), No. e006.

(3) Wu, F.; Robinson, B.; Gao, Y.; Dang, F. New contaminants: existence and knowledge gaps. *New Contaminants* **2025**, *1* (1), No. e001.

(4) Furxhi, I.; Murphy, F.; Mullins, M.; Arvanitis, A.; Poland, C. A. Practices and Trends of Machine Learning Application in Nanotoxicology. *Nanomaterials* **2020**, *10* (1), 116.

(5) Puzyn, T.; Rasulev, B.; Gajewicz, A.; Hu, X.; Dasari, T. P.; Michalkova, A.; Hwang, H.-M.; Toropov, A.; Leszczynska, D.; Leszczynski, J. Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nat. Nanotechnol.* **2011**, *6* (3), 175−178.

(6) Singh, A. V.; Ansari, M. H. D.; Rosenkranz, D.; Maharjan, R. S.; Kriegel, F. L.; Gandhi, K.; Kanase, A.; Singh, R.; Laux, P.; Luch, A.

Downloaded from pubs.acs.org/estlcu/article-pdf/13/8/1114/6625613/acs.estlett.6c00579.pdf by BEIHANG UNIV user on 18 August 2026

Artificial Intelligence and Machine Learning in Computational Nanotoxicology: Unlocking and Empowering Nanomedicine. *Adv. Healthcare Mater.* **2020**, *9* (17). DOI: 10.1002/adhm.201901862.
(7) Yan, X. L.; Yue, T. T.; Winkler, D. A.; Yin, Y. G.; Zhu, H.; Jiang, G. B.; Yan, B. Converting Nanotoxicity Data to Information Using Artificial Intelligence and Simulation. *Chem. Rev.* **2023**, *123* (13), 8575−8637.
(8) Mu, Y.; Wu, F.; Zhao, Q.; Ji, R.; Qie, Y.; Zhou, Y.; Hu, Y.; Pang, C.; Hristozov, D.; Giesy, J. P.; et al. Predicting toxic potencies of metal oxide nanoparticles by means of nano-QSARs. *Nanotoxicology* **2016**, *10* (9), 1207−1214.
(9) Connell, D. W.; Yu, Q. J.; Verma, V. Influence of exposure time on toxicity—An overview. *Toxicology* **2016**, *355*−*356*, 49−53.
(10) Yang, S. P.; Bar-Ilan, O.; Peterson, R. E.; Heideman, W.; Hamers, R. J.; Pedersen, J. A. Influence of Humic Acid on Titanium Dioxide Nanoparticle Toxicity to Developing Zebrafish. *Environ. Sci. Technol.* **2013**, *47* (9), 4718−4725.
(11) Feng, S.; Zhu, L.; Zhao, X.; Sui, Q.; Sun, X.; Chen, B.; Qu, K.; Xia, B. Ecological risk assessment of metallic nanoparticles on the marine environments: Species sensitivity distributions analysis. *Front. Mar. Sci.* **2022**, *9*. DOI: 10.3389/fmars.2022.985195.
(12) Rubach, M. N.; Baird, D. J.; Boerwinkel, M.-C.; Maund, S. J.; Roessink, I.; Van den Brink, P. J. Species traits as predictors for intrinsic sensitivity of aquatic invertebrates to the insecticide chlorpyrifos. *Ecotoxicology* **2012**, *21* (7), 2088−2101.
(13) Zhou, Y.; Wang, Y.; Peijnenburg, W.; Vijver, M. G.; Balraadjsing, S.; Dong, Z.; Zhao, X.; Leung, K. M. Y.; Mortensen, H. M.; Wang, Z.; et al. Application of Machine Learning in Nanotoxicology: A Critical Review and Perspective. *Environ. Sci. Technol.* **2024**, DOI: 10.1021/acs.est.4c03328.
(14) Choi, J.-S.; Ha, M. K.; Trinh, T. X.; Yoon, T. H.; Byun, H.-G. Towards a generalized toxicity prediction model for oxide nanomaterials using integrated data from different sources. *Sci. Rep.* **2018**, *8* (1), 1−10.
(15) Ma, Y.; Wang, J.; Wu, J.; Tong, C.; Zhang, T. Meta-analysis of cellular toxicity for graphene via data-mining the literature and machine learning. *Science of The Total Environment* **2021**, *793*, 148532.
(16) Trinh, T. X.; Ha, M. K.; Choi, J. S.; Byun, H. G.; Yoon, T. H. Curation of datasets, assessment of their quality and completeness, and nanoSAR classification model development for metallic nanoparticles. *Environmental Science-Nano* **2018**, *5* (8), 1902−1910.
(17) Romeo, D.; Louka, C.; Gudino, B.; Wigstrom, J.; Wick, P. Structure-activity relationship of graphene-related materials: A meta-analysis based on mammalian in vitro toxicity data. *Nanoimpact* **2022**, *28*, 100436.
(18) Zhou, Y.; Wang, Y.; Peijnenburg, W.; Vijver, M. G.; Balraadjsing, S.; Fan, W. Using Machine Learning to Predict Adverse Effects of Metallic Nanomaterials to Various Aquatic Organisms. *Environ. Sci. Technol.* **2023**, *57*, 17786.
(19) Balraadjsing, S.; Peijnenburg, W. J. G. M.; Vijver, M. G. Building species trait-specific nano-QSARs: Model stacking, navigating model uncertainties and limitations, and the effect of dataset size. *Environ. Int.* **2024**, *188*, 108764.
(20) Nigro, M.; Bernardeschi, M.; Costagliola, D.; Della Torre, C.; Frenzilli, G.; Guidi, P.; Lucchesi, P.; Mottola, F.; Santonastaso, M.; Scarcelli, V.; et al. n-TiO2 and CdCl2 co-exposure to titanium dioxide nanoparticles and cadmium: Genomic, DNA and chromosomal damage evaluation in the marine fish European sea bass (Dicentrarchus labrax). *Aquatic Toxicology* **2015**, *168*, 72−77.
(21) Thomas, P. D.; Ebert, D.; Muruganujan, A.; Mushayahama, T.; Albou, L.-P.; Mi, H. PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci.* **2022**, *31* (1), 8−22.
(22) Van den Berg, S. J. P.; Baveco, H.; Butler, E.; De Laender, F.; Focks, A.; Franco, A.; Rendal, C.; Van den Brink, P. J. Modeling the Sensitivity of Aquatic Macroinvertebrates to Chemicals Using Traits. *Environ. Sci. Technol.* **2019**, *53* (10), 6025−6034.
(23) Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; et al. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **2000**, *25* (1), 25−29.
(24) Sayers, E. W.; Bolton, E. E.; Brister, J. R.; Canese, K.; Chan, J.; Comeau, D. C.; Connor, R.; Funk, K.; Kelly, C.; Kim, S.; et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **2022**, *50* (D1), D20−d26.
(25) Tang, H.; Finn, R. D.; Thomas, P. D. TreeGrafter: phylogenetic tree-based annotation of proteins with Gene Ontology terms and other annotations. *Bioinformatics* **2019**, *35* (3), 518−520.
(26) Juganson, K.; Ivask, A.; Blinova, I.; Mortimer, M.; Kahru, A. NanoE-Tox: New and in-depth database concerning ecotoxicity of nanomaterials. *Beilstein J. Nanotechnol* **2015**, *6*, 1788−1804.
(27) Chen, G.; Peijnenburg, W. J. G. M.; Xiao, Y.; Vijver, M. G. Developing species sensitivity distributions for metallic nanomaterials considering the characteristics of nanomaterials, experimental conditions, and different types of endpoints. *Food Chem. Toxicol.* **2018**, *112*, 563−570.
(28) Bunmahotama, W.; Vijver, M. G.; Peijnenburg, W. Development of a Quasi-Quantitative Structure-Activity Relationship Model for Prediction of the Immobilization Response of Daphnia magna Exposed to Metal-Based Nanomaterials. *Environ. Toxicol. Chem.* **2022**, *41* (6), 1439−1450.
(29) Ribeiro, M. T.; Singh, S.; Guestrin, C. "Why Should I Trust You?" Explaining the Predictions of Any Classifier. *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, San Francisco, CA, Aug 13−17, 2016; ACM, 2016; pp 1135−1144. DOI: 10.1145/2939672.2939778.
(30) Ankley, G. T.; Bennett, R. S.; Erickson, R. J.; Hoff, D. J.; Hornung, M. W.; Johnson, R. D.; Mount, D. R.; Nichols, J. W.; Russom, C. L.; Schmieder, P. K.; et al. Adverse Outcome Pathways: A Conceptual Framework to Support Ecotoxicology Research and Risk Assessment. *Environ. Toxicol. Chem.* **2009**, *29* (3), 730−741.
(31) Abdolahpur Monikh, F.; Chupani, L.; Guo, Z.; Zhang, P.; Darbha, G. K.; Vijver, M. G.; Valsami-Jones, E.; Peijnenburg, W. The stochastic association of nanoparticles with algae at the cellular level: Effects of NOM, particle size and particle shape. *Ecotoxicol Environ. Saf* **2021**, *218*, 112280.
(32) Wang, D.; Lin, Z.; Wang, T.; Yao, Z.; Qin, M.; Zheng, S.; Lu, W. Where does the toxicity of metal oxide nanoparticles come from: The nanoparticles, the ions, or a combination of both? *Journal of Hazardous Materials* **2016**, *308*, 328−334.
(33) Kar, S.; Pathakoti, K.; Tchounwou, P. B.; Leszczynska, D.; Leszczynski, J. Evaluating the cytotoxicity of a large pool of metal oxide nanoparticles to Escherichia coli: Mechanistic understanding through In Vitro and In Silico studies. *Chemosphere* **2021**, *264*, 128428.
(34) Wang, W.; Gaus, K.; Tilley, R. D.; Gooding, J. J. The impact of nanoparticle shape on cellular internalisation and transport: what do the different analysis methods tell us? *Materials Horizons* **2019**, *6* (8), 1538−1547.
(35) Liu, R.; Zhang, H. Y.; Ji, Z. X.; Rallo, R.; Xia, T.; Chang, C. H.; Nel, A.; Cohen, Y. Development of structure-activity relationship for metal oxide nanoparticles. *Nanoscale* **2013**, *5* (12), 5644−5653.
(36) Yu, Z.; Li, Q.; Wang, J.; Yu, Y.; Wang, Y.; Zhou, Q.; Li, P. Reactive Oxygen Species-Related Nanoparticle Toxicity in the Biomedical Field. *Nanoscale Res. Lett.* **2020**, *15* (1), 115.
(37) Xiong, T.; Zhang, S.; Kang, Z.; Zhang, T.; Li, S. Dose-Dependent Physiological and Transcriptomic Responses of Lettuce (*Lactuca sativa* L.) to Copper Oxide Nanoparticles-Insights into the Phytotoxicity Mechanisms. *International Journal of Molecular Sciences* **2021**, *22* (7), 3688.
(38) Saulou-Berion, C.; Gonzalez, I.; Enjalbert, B.; Audinot, J.-N.; Fourquaux, I.; Jamme, F.; Cocaign-Bousquet, M.; Mercier-Bonin, M.; Girbal, L. Escherichia coli under Ionic Silver Stress: An Integrative Approach to Explore Transcriptional, Physiological and Biochemical Responses. *PLoS One* **2015**, *10* (12), e0145748.
(39) Zhao, X.; Wang, S.; Wu, Y.; You, H.; Lv, L. Acute ZnO nanoparticles exposure induces developmental toxicity, oxidative

Downloaded from pubs.acs.org/estlcu/article-pdf/13/8/1114/6625613/acs.estlett.6c00579.pdf by BEIHANG UNIV user on 18 August 2026

stress and DNA damage in embryo-larval zebrafish. *Aquatic Toxicology* **2013**, *136*, 49−59.
(40) Lan, X.; Huang, W.; Sun, B.; Waiho, K.; Song, H.; Hu, M.; Khalid, M.; Wang, Y. Combined effects of pentachlorophenol and nano-TiO_2 with different sizes on antioxidant, digestive, and immune responses of the swimming crab Portunus trituberculatus. *Aquatic Toxicology* **2024**, *270*, 106900.
(41) Rallo, R.; France, B.; Liu, R.; Nair, S.; George, S.; Damoiseaux, R.; Giralt, F.; Nel, A.; Bradley, K.; Cohen, Y. Self-organizing map analysis of toxicity-related cell signaling pathways for metal and metal oxide nanoparticles. *Environ. Sci. Technol.* **2011**, *45* (4), 1695−1702.
(42) Ivask, A.; Titma, T.; Visnapuu, M.; Vija, H.; Kakinen, A.; Sihtmae, M.; Pokhrel, S.; Madler, L.; Heinlaan, M.; Kisand, V.; Shimmo, R.; Kahru, A.; et al. Toxicity of 11 Metal Oxide Nanoparticles to Three Mammalian Cell Types In Vitro. *Curr. Top. Med. Chem.* **2015**, *15* (18), 1914−1929.
(43) George, S.; Xia, T.; Rallo, R.; Zhao, Y.; Ji, Z.; Lin, S.; Wang, X.; Zhang, H.; France, B.; Schoenfeld, D.; et al. Use of a High-Throughput Screening Approach Coupled with In Vivo Zebrafish Embryo Screening To Develop Hazard Ranking for Engineered Nanomaterials. *ACS Nano* **2011**, *5* (3), 1805−1817.
(44) Lee, J. W.; Won, E.-J.; Raisuddin, S.; Lee, J.-S. Significance of adverse outcome pathways in biomarker-based environmental risk assessment in aquatic organisms. *J. Environ. Sci.* **2015**, *35*, 115−127.

Downloaded from pubs.acs.org/estlcu/article-pdf/13/8/1114/6625613/acs.estlett.6c00579.pdf by BEIHANG UNIV user on 18 August 2026

https://doi.org/10.1021/acs.estlett.6c00579
Environ. Sci. Technol. Lett. 2026, 13, 1114−1121