PNAS RESEARCH ARTICLE | ECOLOGY

Dark diversity framework reconciles Darwin's naturalization conundrum for freshwater fish invasions

Wen-gang Zhang[a,1], Meng Xu[b,c,1], Shu-ya Fan[a,d,e], Jonathan A. Bennett[f], Ross N. Cuthbert[g], Jaimie T. A. Dick[g], Meelis Pärtel[h], and Shao-peng Li[a,2]

Affiliations are included on p. 8.

Edited by David M. Hillis, The University of Texas at Austin, Austin, TX; received February 18, 2026; accepted May 8, 2026

The spread of invasive species poses a major threat to global biodiversity, yet predicting successful establishment of exotic species in novel environments remains challenging. Darwin's naturalization conundrum is a longstanding debate over whether exotic species closely or distantly related to native communities are more likely to succeed. Despite its long history, empirical studies continue to yield conflicting evidence. Here, we introduce the dark diversity concept, which refers to species that could theoretically inhabit a site but are currently absent. Combined with observed diversity, dark diversity integrates information on the potential diversity (site-specific species pool size) of the resident community and how completely that pool is locally present (community completeness). Analyzing a 340-y record of successful and failed fish introductions across 516 Swedish lakes, we showed that the effect of phylogenetic relatedness on invasion outcome depends on dark diversity context. Exotics closely related to resident species were more likely to establish in communities with smaller species pools and higher completeness, whereas phylogenetically distant exotic species were more successful in communities with larger species pools and lower completeness. Models with observed species richness obtained considerably less support. Thus, integrating the dark diversity framework clarifies the contrasting effects of phylogenetic relatedness on invasion outcomes, and helps reconcile this 160-y-old conundrum.

biological invasions | community assembly | Darwin's naturalization hypothesis | phylogenetic relatedness | species pool

The global spread of invasive species has led to profound ecological and socioeconomic consequences, threatening biodiversity, altering community structures, and modifying ecosystem functions (1–3). Predicting which exotic species are likely to establish in which environments is thus critical and a central goal of invasion science (4–6). This process is inherently complex, as invasion success depends not only on a species' adaptation to local environmental conditions but also on its interactions with native species, as well as propagule pressure and socioeconomic context (7–9). Although generalities remain elusive, increasing attention has focused on how phylogenetic relatedness between exotic and native species jointly shapes invasion outcomes, stemming from principles forged originally by Charles Darwin (10–12).

On one hand, Darwin posited that exotic species closely related to native species are more likely to establish, as they share similar environmental preferences with native species, which is referred to as the preadaptation hypothesis (12, 13). On the other hand, Darwin also suggested that exotic species phylogenetically distant to native species would tend to be more successful, as they may face less competition and share fewer natural enemies with native species in the novel environments (14). This hypothesis is known as Darwin's naturalization hypothesis (15). Together, these two seemingly contradictory hypotheses, which predict whether exotic species that are closely or distantly related to native species are more likely to succeed, constitute Darwin's naturalization conundrum (15, 16). While this conundrum has been widely tested across taxa and systems, accumulating evidence indicates the two hypotheses are not mutually exclusive, but their relative support varies with ecological context (17–20). The preadaptation hypothesis is more likely to be supported in regions characterized by stressful abiotic conditions, where environmental filtering is the primary driver of species establishment (19, 21–23). In contrast, Darwin's naturalization hypothesis is more applicable to local communities in benign environments with high species richness or abundant natural enemies, where niches are already occupied by closely related natives and exotic species face stronger competition and enemy pressure, reducing establishment of closely related exotics (11, 24, 25). However, despite this progress, identifying the ecological contexts in which each hypothesis applies remains difficult,

Significance

Predicting which exotic species will establish in new environments is a central challenge for invasion ecology. Since Darwin first posed his naturalization conundrum, ecologists have long debated whether exotic species similar or dissimilar to the native species are more likely to succeed. We argue that resolving this debate may require considering dark diversity, which comprehensively captures species pool size and community completeness. By incorporating this overlooked dimension, we showed smaller species pools and higher community completeness favor the establishment of exotic fishes closely related to native fishes in the Swedish lakes, whereas larger species pools and lower community completeness favor distant exotics. This dark diversity framework reframes Darwin's conundrum and provides a practical tool for forecasting invasions and guiding conservation.

Author contributions: S.-P.L. designed research; W.-G.Z. and M.X. performed research; R.N.C., J.T.A.D., M.P., and S.-P.L. contributed new reagents/analytic tools; W.-G.Z., M.X., S.-Y.F., and M.P. analyzed data; and W.-G.Z., M.X., S.-Y.F., J.A.B., R.N.C., J.T.A.D., M.P., and S.-P.L. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1] W.-G.Z. and M.X. contributed equally to this work.

[2] To whom correspondence may be addressed. Email: spli@des.ecnu.edu.cn.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2604929123/-/DCSupplemental.

as the environmental and biotic factors influencing species establishment are often difficult to quantify and disentangle.

We propose that the dark diversity framework, incorporating site-specific species pool size and community completeness, provides a perspective for understanding the context dependence of Darwin's naturalization conundrum (Fig. 1). Dark diversity refers to species that are currently absent from a given site but could theoretically inhabit it based on local ecological conditions (26, 27). These absent species, along with the observed local diversity, form the site-specific species pool, which reflects the biodiversity potential of a local community (Fig. 1*A*). Consequently, smaller species pool sizes often reflect restrictive abiotic filters (28, 29), such as low dissolved oxygen or resource-limited in aquatic habitats, which may only allow the establishment of closely related species with similar ecological requirements, consistent with the preadaptation hypothesis (Fig. 1 *B* and *D*). In contrast, larger species pools, indicative of more permissive environments and diverse niches, may enable the establishment of distantly related species to exploit distinct ecological roles, in support of Darwin's naturalization hypothesis. Community completeness, defined as the ratio of observed to dark diversity, represents the extent to which the species pool is realized in a given community (Fig. 1*A*) (27, 29). Low community completeness, characterized by a large proportion of excluded potential species, may indicate unfilled niches and be often associated with strong competitive exclusion (30, 31). Such conditions may favor the establishment of phylogenetically distant species that occupy novel niches with minimal direct competition (32–34), thus supporting Darwin's naturalization hypothesis (Fig. 1 *C* and *D*). Conversely, high-completeness suggests that niches are largely occupied, with few species excluded by biotic interactions. In these saturated communities, any new species must fit into an already crowded niche space. Consequently, invaders preadapted to local conditions, often close relatives of native species, are most likely to establish, sometimes by outcompeting natives, consistent with the preadaptation hypothesis (11, 35–37). Importantly, both species pool size and community completeness can be readily estimated from native community data (38–40), which act as indirect surrogates for niche diversity and filling. Therefore, the dark diversity framework provides a practical perspective for reconciling Darwin's naturalization conundrum, particularly in contexts where direct measurements of complex environmental variables and species interactions are unavailable or impractical.

Freshwater fishes are among the most frequently introduced animal taxa, with human activities such as aquaculture, recreational trade, and ballast water transport facilitating their movement far beyond their native ranges (5, 41). Along with frequent passive dispersal, at least 551 exotic fish species have established self-sustaining populations in more than half of the world's freshwater basins, posing a significant threat to these ecosystems (35, 42). Most empirical studies on freshwater fishes emphasize the dominant role of environmental preadaptation in establishment and the higher success rates of closely related exotic species (35, 43, 44). In contrast, the role of biotic interactions remains underexplored, though some studies have emphasized their importance and reported greater success of distantly related species (5, 35, 45, 46). To explain these

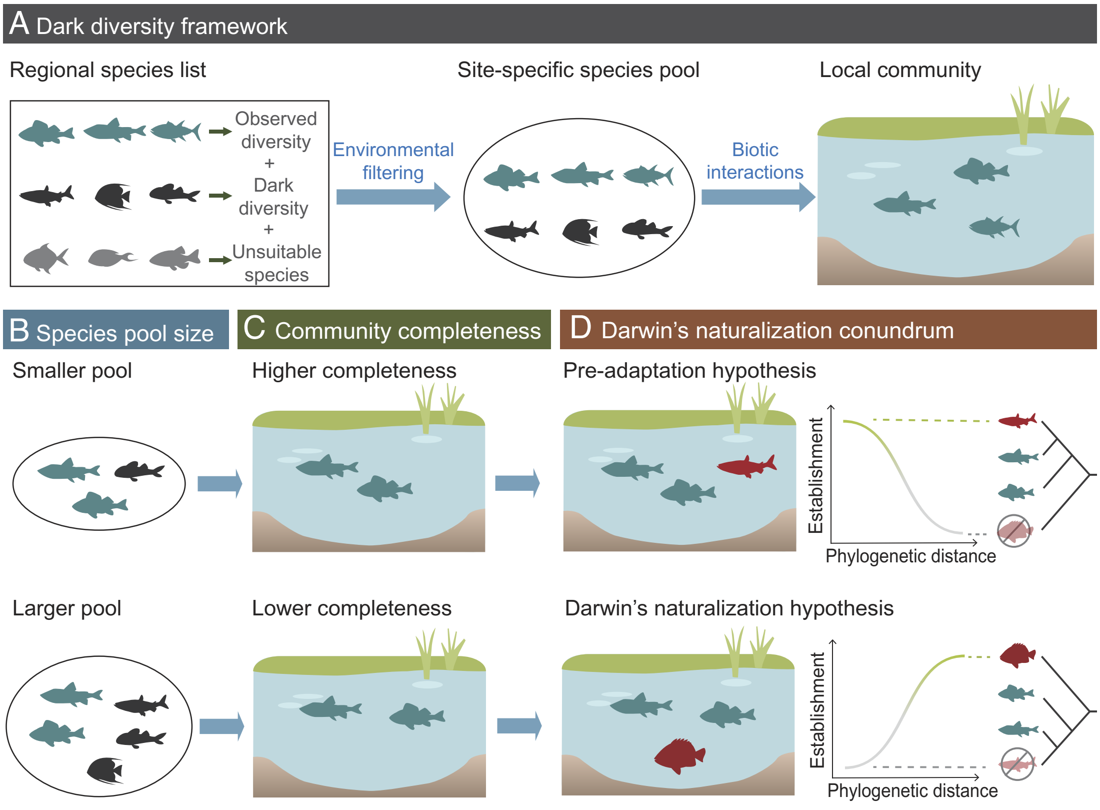

Fig. 1. The dark diversity concept as a framework for reconciling Darwin's naturalization conundrum. (*A*) Dark diversity (black) refers to species absent from a local community but capable of inhabiting it, together with observed diversity (blue) it defines the site-specific species pool and community completeness (ratio of observed to dark diversity). (*B*) Smaller species pools, likely reflecting stronger abiotic filtering, would favor the establishment of exotics (red) that are closely related to native species, supporting the preadaptation hypothesis (PAH). In contrast, larger species pools may facilitate the success of distantly related exotics (red), aligning with Darwin's naturalization hypothesis (DNH). (*C*) Lower community completeness, characterized by more unfilled niches due to biotic exclusion of native species, favors the establishment of distantly related species, consistent with DNH. Conversely, higher community completeness, indicating fewer unfilled niches and reduced biotic interactions, enhances the success of closely related species, supporting PAH. (*D*) The PAH predicts exotic species phylogenetically close to native species are more likely to establish, whereas DNH predicts exotic species distantly related to natives are more likely to succeed.

contrasting outcomes, we applied the dark diversity framework to investigate the phylogenetic patterns of fish establishment across 516 freshwater lakes in Sweden, using a 340-y dataset documenting 748 introduction events of 22 exotic fish species (Fig. 2 and *SI Appendix*, Fig. S1) (47). This unique dataset, which documents initial community compositions and the outcomes of each introduction event, provides a valuable opportunity to empirically test Darwin's naturalization conundrum. Using these data, we estimated the dark diversity of each community with the hypergeometric method (39), and subsequently calculated site-specific species pool size and community completeness (Fig. 2). For each introduction record, we performed a species-level analysis to relate establishment probability to the phylogenetic distance between each introduced fish and the resident assemblage (including both native and previously established exotic species), and how this relationship is modulated by species pool size and community completeness. For each lake, we conducted a community level analysis to assess how species pool size and community completeness modulate the phylogenetic distances between successfully established exotic species and the native species in the lake. Our findings demonstrate that the validity of preadaptation hypothesis and Darwin's naturalization hypothesis is contingent on both the site-specific species pool size and community completeness of the native communities, highlighting the potential of the dark diversity framework as a unified approach for reconciling Darwin's naturalization conundrum.

Results

Phylogenetic Patterns of Fish Establishment at Species Level. At the species level, the effect of phylogenetic distance to the resident assemblage on the establishment success of introduced fishes depended strongly on site-specific species pool size and community completeness (Fig. 3 and *SI Appendix*, Fig. S2 and Tables S1–S4). Across all introduction records, there was a significant negative effect of both mean phylogenetic distance (MPD) and phylogenetic nearest-neighbor distance (PNND) between exotic and resident species on establishment probability ($P = 0.002$, 0.001, respectively; Fig. 3*A* and *SI Appendix*, Fig. S2 and Tables S1–S4), indicating that introduced fishes more closely related to resident species were more likely to establish, supporting preadaptation hypothesis. Additionally, species pool size had a significant negative effect ($P = 0.014$), while community completeness had a positive effect on establishment success ($P = 0.004$; Fig. 3*A* and *SI Appendix*, Fig. S1 and Tables S1–S4), indicating that exotic fishes were more likely to establish in lakes with smaller species pools and higher community completeness. Further analysis revealed a significant positive interaction between exotic-resident phylogenetic distance (MPD and PNND) and species pool size in explaining establishment probability (Fig. 3*A* and *SI Appendix*, Fig. S2*A*). In lakes with smaller species pools, exotic fishes more closely related to the resident community had a higher probability of establishment, supporting the preadaptation hypothesis. However, as species pool size increased, this pattern reversed, and distantly related species became more likely to establish, aligning with Darwin's naturalization hypothesis (Fig. 3*B* and *SI Appendix*, Fig. S2*B*). We also observed a significant negative interaction between exotic-resident phylogenetic distance (MPD and PNND) and community completeness (Fig. 3*A* and *SI Appendix*, Fig. S2*A*). In lakes with higher completeness, closely related species were more likely to establish, supporting preadaptation hypothesis. In contrast, as completeness decreased, distantly related species had a higher probability of establishment, consistent with Darwin's naturalization hypothesis (Fig. 3*C* and *SI Appendix*, Fig. S2*C*). In contrast, without considering the dark diversity framework, species richness alone had neither a significant direct effect nor an interaction with phylogenetic distance in explaining establishment outcomes, and models using richness as predictor had both lower explanatory power and model

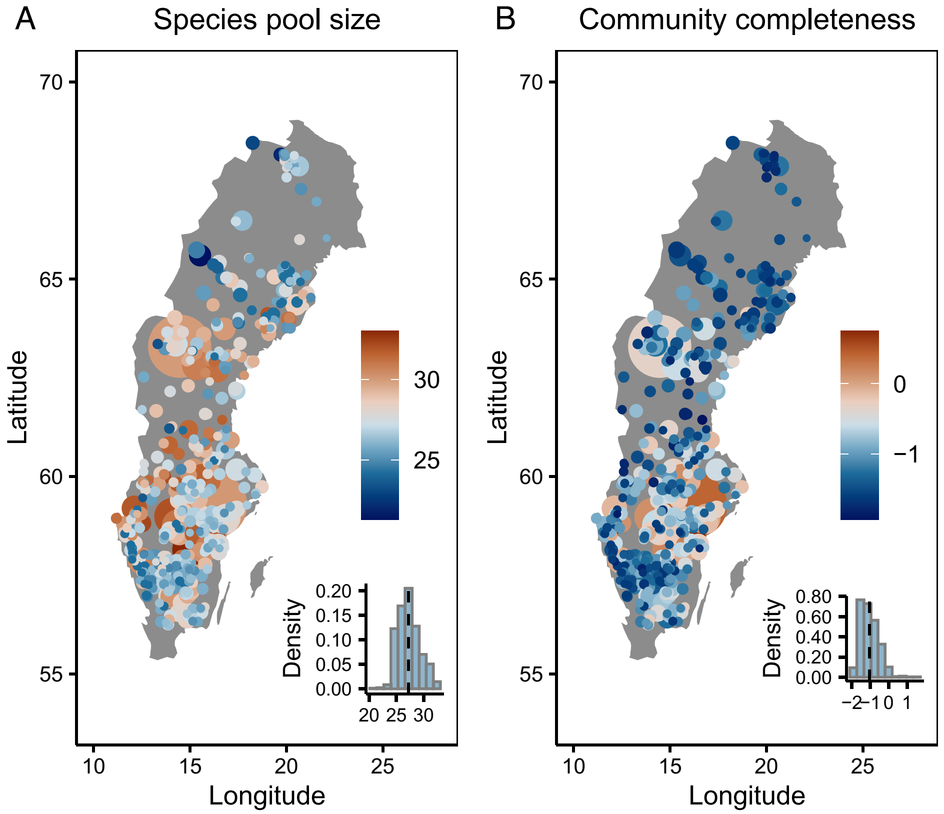

Fig. 2. Species pool size (*A*), community completeness (*B*), and geographical distribution of 516 studied freshwater lakes in Sweden. Circle size corresponds to lake area, with the color gradient indicating the site-specific species pool size (*A*) and community completeness (*B*) for each lake. Community completeness was calculated as ln (observed/dark diversity). Histograms display the density distribution of species pool size (*A*) and community completeness (*B*) across all lakes, with the black dashed line showing the mean value. Map lines delineate the study area and are not necessarily national boundaries.

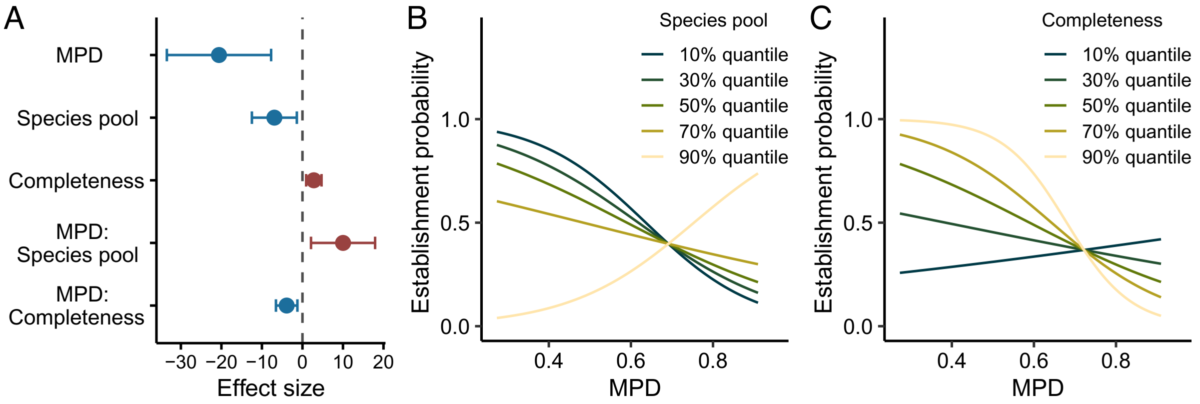

Fig. 3. The effect of phylogenetic distance on the establishment of exotic fishes varies with species pool size and community completeness. (*A*) The effect sizes of mean phylogenetic distance (MPD) between exotic and resident species, species pool size, community completeness, and their interactions on the establishment probability of exotic fishes (n = 748 introduction events). Circles indicate coefficient estimates, with error bars representing 95% confidence intervals. Blue circles denote negative effects, while red circles indicate positive effects. See *SI Appendix*, Tables S1 for further details. (*B* and *C*) Model predictions showing how the relationship between MPD and establishment probabilities of exotic fishes changes with species pool size (*B*) and community completeness (*C*), fixed at the 10%, 30%, 50%, 70%, and 90% quantiles. Species pool size was reflected-log transformed prior to model fitting. All statistical analyses and prediction curves were generated using generalized linear mixed models (GLMMs) with a binomial error distribution.

support according to Akaike Information Criterion (AIC) and Akaike weights (AW) (*SI Appendix*, Fig. S3 and Tables S1 and S3).

These patterns were robust to alternative definitions and estimation methods of dark diversity. Specifically, results were consistent when using a stricter definition of the potential pool (suitability > 0.7), or when restricting the regional species list to species recorded within the same biogeographic region (Fig. 3 and *SI Appendix*, Figs. S4 and S5). We also obtained similar patterns when estimating dark diversity using the Beals-based favorability method instead of the hypergeometric co-occurrence approach (*SI Appendix*, Fig. S6). The results remained consistent when adding lake area as a covariate (*SI Appendix*, Fig. S7). To minimize potential effects of serial introductions, we repeated the analysis using only the first recorded introduction per lake and defining the resident assemblage as the native community prior to any introductions, and the results were similar (*SI Appendix*, Fig. S8). Finally, because phylogenetic distance can covary with resident species richness, we performed null models that held resident richness constant while randomizing either resident community composition or the identity of the introduced species. Standardized effect sizes (SES) of phylogenetic distances from these null models yielded results similar to those based on raw phylogenetic distances (*SI Appendix*, Fig. S9).

Phylogenetic Patterns of Fish Establishment at Community Level. At the community level, we considered all introduction records together, and calculated the phylogenetic distance between all successfully established exotic fishes and the initial native assemblage in each lake. We found the exotic-native phylogenetic distance was significantly influenced by the site-specific species pool size and community completeness of each lake (Fig. 4 and *SI Appendix*, Fig. S10 and Tables S5 and S6). Importantly, species pool size and community completeness were better predictors of exotic-native phylogenetic distance compared to other lake characteristics, including lake area, latitude, elevation, and accumulated water temperature (Fig. 4*A* and *SI Appendix*, Fig. S10 and Tables S5 and S6). Specifically, both the mean pairwise phylogenetic distance (MPPD) and mean nearest taxon distance (MNTD) between

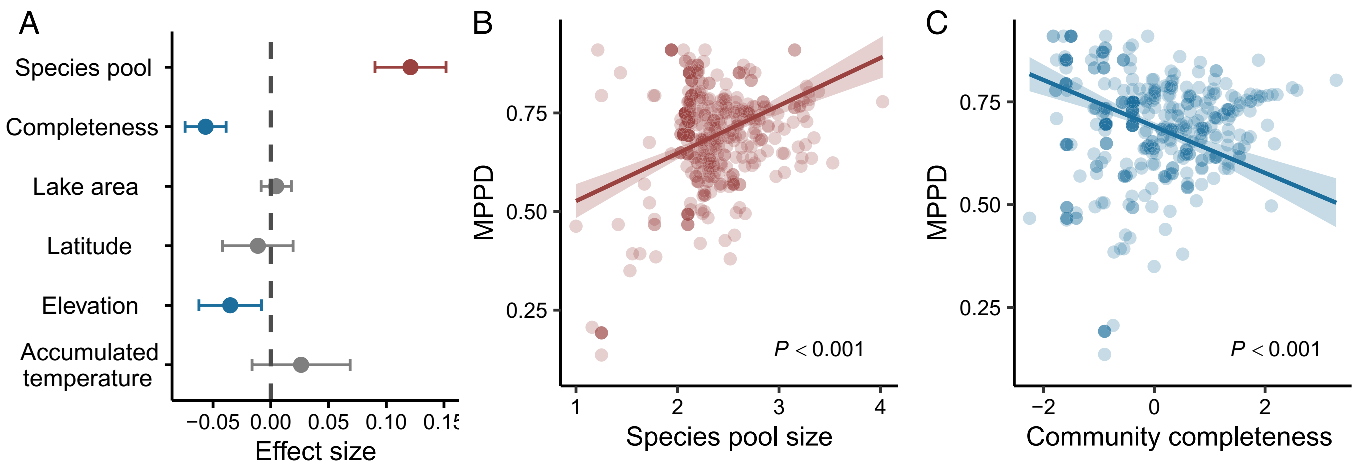

Fig. 4. Effects of species pool size, community completeness, and environmental conditions on phylogenetic distance between established exotic and native fishes. (*A*) The effect sizes of different abiotic and biotic variables on the MPPD between all successfully established exotic and the initial native fishes in each lake (n = 516 lakes). Circles represent standardized effect sizes, with error bars indicating 95% confidence intervals. Blue circles represent negative effects, red circles represent positive effects, while gray circles indicate nonsignificant effects. See *SI Appendix*, Tables S5 for further details. (*B* and *C*) The correlations of exotic-native MPPD with species pool size (*B*) and community completeness (*C*). Solid lines illustrate significant effects at the $\alpha = 0.05$ level, with shaded areas indicating 95% confidence intervals. Species pool size was reflected-log transformed prior to model fitting. All statistical tests and prediction curves were obtained using linear models.

all successfully established exotic and native fishes increased with species pool size ($P < 0.001$, $P = 0.003$, respectively; Fig. 4*B* and *SI Appendix*, Fig. S10*B*), indicating that larger species pools facilitated the establishment of distantly related exotic fishes, in line with Darwin's naturalization hypothesis. In contrast, both MPPD and PNND significantly decreased with increasing community completeness ($P < 0.001$; Fig. 4*C* and *SI Appendix*, Fig. S10*C*), showing that successful exotic species tend to be more closely related to natives in lakes with higher completeness, consistent with the preadaptation hypothesis. Furthermore, models that used species pool size and community completeness as predictors of exotic–native phylogenetic distance consistently had higher explanatory power and stronger support than models based on observed species richness (AW ≈ 1 and ≈ 0, respectively; Fig. 4 and *SI Appendix*, Figs. S10 and S11 and Tables S5 and S6).

To assess the robustness of community-level patterns, we also conducted a series of robustness analyses. We found consistent results when we 1) estimated dark diversity using a more conservative probability threshold (>0.7) (*SI Appendix*, Figs. S12), 2) restricted the regional species list to those recorded within the same biogeographic region (*SI Appendix*, Figs. S13), 3) estimated dark diversity using the Beals-based favorability method instead of the hypergeometric co-occurrence approach (*SI Appendix*, Fig. S14), and 4) defined native communities using the most recent native species records rather than the initial assemblage (*SI Appendix*, Fig. S15). Further, null-model standardized distances (i.e., SES.MPPD and SES.MNTD), generated by randomizing either the native or the exotic assemblage while holding richness unchanged, yielded patterns consistent with the observed distances (*SI Appendix*, Fig. S16).

Discussion

As a longstanding paradox in invasion ecology, Darwin's naturalization conundrum has gained renewed attention over the past three decades, particularly in the context of fish invasions (12, 14, 15, 20, 43). Although most studies of freshwater fish support the preadaptation hypothesis (35, 43, 44), substantial evidence also supports Darwin's naturalization hypothesis (5, 45, 46), making it difficult to specify a priori which hypothesis will apply in a given community. In this study, we provided a lens for reconciling this conundrum by employing the dark diversity framework, which integrates site-specific species pool sizes and community completeness to capture both environmental and biotic effects. We revealed that the preadaptation hypothesis is more relevant in communities with smaller species pools and higher completeness, while Darwin's naturalization hypothesis is more likely supported in communities with larger species pools and lower completeness. The models with observed species richness received much less support. These patterns were robust to alternative phylogenetic metrics, including mean and nearest phylogenetic distances, which capture overall vs. nearest phylogenetic relatedness to recipient communities and reflect diffuse biotic resistance from all native species versus stronger effects of closest relatives, respectively (11, 15, 16). Our results are consistent with recent findings showing that the preadaptation hypothesis is better supported in constrained environments where restrictive abiotic filters limit establishment, but that Darwin's naturalization hypothesis is more valid in less constrained environments with greater biotic interactions and available niches (19, 21, 23, 24). Moreover, species pool size and community completeness, as central elements of the dark diversity framework, can be estimated based on native community data, allowing applications even in contexts where direct measurements of complex environmental variables and species interactions are not feasible (38–40). Furthermore, we demonstrated that the dark diversity framework explained the context dependence of Darwin's naturalization conundrum more effectively than traditionally considered factors like species richness, lake area, elevation, latitude, and water temperature. Together, these findings underscore the value of quantifying species pool size and community completeness for better understanding the establishment of exotic species.

The preadaptation hypothesis, which posits that closely related exotic species are more likely to establish, has long been considered a consequence of environmental filtering. Harsh environmental conditions may limit community composition to a few species with similar adaptive traits, favoring the establishment of closely related exotic species that share these requirements with native species (13, 16, 36). Numerous studies have explored Darwin's naturalization conundrum along environmental gradients, such as elevation and latitude (21, 23), and in relation to specific environmental variables like temperature and precipitation (20, 24). These studies often found stronger evidence in supporting preadaptation hypothesis in more constrained environments (18, 19, 22). However, revealing the roles of specific environmental variables remains challenging due to their complex interactions, and their relative importance in determining environmental filtering is often indistinguishable (6, 7, 24). In contrast, species pool size provides a more integrated measure, capturing both measured and unmeasured abiotic variables that influence environmental constraints (28, 29, 48). In this study, our findings suggested that, for freshwater fish introductions in Swedish lakes, species pool size may serve as a more informative predictor of the phylogenetic patterns of species establishment than specific environmental variables such as lake area, latitude, elevation, or water temperature (Fig. 4*A* and *SI Appendix*, Fig. S10*A*). Moreover, smaller species pools may reflect not only specific environmental filters but also broader ecological processes that constrain species establishment, such as stronger disturbances or environmental fluctuations, which may further favor the establishment of closely related species (17, 21, 49, 50). Together, our study demonstrates that species pool size can serve as a holistic indicator of site-level environmental suitability, capturing a wide range of ecological processes that influence exotic species establishment.

In contrast to the preadaptation hypothesis, Darwin's naturalization hypothesis has long been attributed to biotic interactions, where distantly related species are thought to avoid competition by occupying novel ecological roles (16, 19, 34) and escaping from natural enemies that target closely related species (37, 51). However, quantifying the strength and ecological consequences of such interactions remains inherently challenging, especially in fish communities (12, 16, 52). Owing to their mobility, fish may mitigate interspecific pressures through movement, and their interactions often involve complex dynamics such as intraguild predation, hybridization, and cross-trophic effects (35, 53, 54). Nevertheless, despite being poorly understood, biotic interactions are likely to exert meaningful influences on species establishment in spatially and resource-limited systems such as lakes (45, 55–57). Besides, if there are more underutilized resources and unfilled niches within the community, this may also facilitate the establishment of phylogenetically distant exotic species that occupy distinct niches from natives (32–34). Our findings suggest that community completeness, by quantifying the realization ratio of the species pool, provides a suitable proxy for the strength of biotic interactions and the extent of niche filling in these freshwater fish communities (26, 29, 58). In particular, lower community completeness would indicate stronger biotic interactions or more unfilled niches (26, 29, 58), hindering the establishment of closely related exotic species and aligning with the prediction of Darwin's naturalization hypothesis. Importantly, our results showed that community completeness was a more effective predictor of the validity of Darwin's naturalization hypothesis than species richness. While species richness is a commonly used proxy for local biotic interactions (e.g., competition), it also depends on the potential diversity that communities can support, which may limit its effectiveness (52, 59, 60). In contrast, by incorporating dark diversity, community completeness isolates the influence of biotic interactions on the site-specific species pool from the confounding effects of environmental variability (29, 58). Because it is defined relative to each community's species pool, completeness also enables meaningful comparisons of biotic saturation across heterogeneous and different regions. In addition, lower completeness may also arise from nonbiotic processes such as dispersal limitation or stochastic variation (29, 31), which restrict species' access to available niches and may similarly favor the establishment of distantly related species. Thus, our study supports the idea that community completeness could integrate complex biotic and niche-based processes into a single and informative metric, highlighting the potential of the dark diversity framework to enhance predictions of invasion outcomes when disentangling each ecological interaction individually is not feasible.

This study provides an application of the dark diversity concept to invasion biology, demonstrating its capacity to reconcile the context dependence of Darwin's naturalization conundrum. However, several limitations should be acknowledged. First, while Darwin's naturalization conundrum research has expanded to include taxa such as plants, birds, fishes, and microorganisms (12, 13, 61, 62), our analysis was restricted to a single taxonomic group (freshwater fish) in one region (Swedish lakes). Therefore, testing the dark diversity framework across additional taxa will be crucial to assessing its generality and potentially for resolving this 160-y-old conundrum. Second, the lack of data on the number of introduced individuals and native species may introduce uncertainty, as propagule pressure and numerical dominance are known to influence establishment success (5, 6, 35). Third, our study primarily focuses on regional scale and establishment stage of fish invasions. Given that the phylogenetic patterns of invasion often vary across spatial scales, modes of introduction, and invasion stages (11, 22, 36, 37, 62), future research should explore the application of the dark diversity framework across multiple scales and at other stages of invasion. Fourth, while recent Darwin's naturalization conundrum research has increasingly incorporated multidimensional ecological differences, such as phylogenetic, functional, niche, and fitness differences (16, 17, 61, 63), it remains critical to assess whether the dark diversity framework can resolve the inconsistencies generated by these diverse difference measures. However, quantifying ecological differences beyond phylogenetics in long-term observational datasets remains particularly challenging due to their inherent variability. Finally, by definition, dark diversity cannot be directly observed and must be inferred statistically (40). While previous work has provided consistent support for the robustness of hypergeometric co-occurrence–based methods across diverse ecological contexts (27, 38–40), they still rely on underlying assumptions about species associations and data structure of the available data, and co-occurrence patterns provide only simplified summaries of underlying ecological relationships (39). Ongoing methodological advances are likely to further improve dark diversity inference and expand the scope of this framework.

In summary, our study highlights the value of the dark diversity framework in advancing our understanding of Darwin's naturalization conundrum. Specifically, we suggest that the estimation of species pool size and community completeness based on native community data offers practical, integrated metrics of local environmental suitability and biotic interactions, moving holistically beyond species richness and specific environmental factors. This dark diversity framework disregards detailed ecological processes but instead provides two convenient indicators for predicting the conditions under which the preadaptation hypothesis or Darwin's naturalization hypothesis is more likely to be supported. Moving forward, we encourage future research to apply the dark diversity framework across a broader range of taxa, habitats, and community types, considering different invasion stages and spatial scales. Such studies will deepen our understanding of how ecological differences between exotic and native species influence biological invasion across various scenarios.

Materials and Methods

Data Collection. We analyzed a comprehensive dataset of 1,157 documented introductions of 26 freshwater fish species across 821 Swedish lakes during 1658–2002, as sourced from Henriksson et al. (47). This dataset meticulously compiled both the native species composition of the lakes and the outcomes of introductions (failure or success) for each introduction record. Successful establishment was defined as the introduction of a fish that either reproduced in the lake or remained present for more than 20 y (47). Based on efforts to compile a broader national database, these records integrate historical reports, gillnet data, and interviews with local fishermen, and were rigorously validated to ensure data quality (47, 64). For each lake, area, latitude, longitude, and elevation were recorded. Accumulated water temperature was directly measured for 198 lakes, while for the remaining lakes it was estimated using a multiple regression function based on measured data and geographic characteristics (47, 63). To ensure accurate calculation of phylogenetic distances, we followed the approach of Xu et al. (63) by further refining the dataset to include 965 introduction records of species with confirmed presence and clear phylogenetic information. Moreover, we only focused on lakes with at least three species to ensure a sufficient number of observations for accurately constructing species co-occurrence matrices and estimating dark diversity (38, 39). As a result, our analysis was conducted using a refined dataset comprising 748 introduction records for 22 fishes across 516 lakes (Fig. 2 and *SI Appendix*, Fig. S1). Among these records, 337 introductions (45%) were successful.

Phylogenetic Analyses. We obtained molecular phylogenetic data for all fish species from Xu et al. (63). Briefly, Bayesian phylogenetic trees for these species were constructed based on three mitochondrial gene sequences (*cytb*, *COI*, and *16S rRNA*) retrieved from GenBank, with *Carcharodon carcharias* serving as an outgroup. Subsequently, we calculated the pairwise phylogenetic distances between all species in the dataset using the *cophenetic.phylo* function of the R package ape (65).

At the species level, for each introduction record, we classified all fish species present prior to the introduction as resident assemblage, while the introduced species was categorized as exotic (47, 63, 64). Accordingly, in lakes with multiple introduction records, previously established exotic species were treated as part of the resident community for each introduction event, as they coexist with native species and jointly influence the establishment of subsequent introductions. To avoid potential influence from serial introductions, we repeated the analysis using only the first recorded introduction per lake, and defining the resident assemblage as the native community prior to any introductions, which produced results consistent with those from the full dataset (Fig. 3 and *SI Appendix*, Fig. S8). We calculated two metrics using the R package picante (66) to assess the phylogenetic relationship between each exotic species and the resident communities: the mean phylogenetic distance (MPD) and the phylogenetic nearest neighbor distance (PNND) between each exotic fish and all resident fish species in the lake (16, 67, 68). MPD quantifies the average phylogenetic relatedness between an exotic species and the entire resident assemblage, thereby capturing deeper patterns in the phylogeny. In contrast, PNND reflects the distance to the closest relative and is more sensitive to terminal phylogenetic relationships. Together, these metrics provide complementary information, with MPD representing the overall phylogenetic distinctiveness of the invader, whereas PNND captures the potential for strong niche overlap with the most similar resident species (11, 16). To assess whether the effects of observed phylogenetic distance on establishment differ from random assembly while controlling for species richness, we computed standardized effect sizes (SES) of MPD and PNND using two complementary null models (4, 11, 15). In the first null model, we randomized the resident assemblage by drawing species without replacement from the full resident species pool, while keeping resident richness and the identity of the introduced species fixed. In the second null model, we randomized the introduced species identity by sampling from all exotic species in the dataset, while keeping the resident richness and composition fixed. Each null model was run 999 times to generate null distributions of MPD and PNND, from which SES values were calculated as the difference between the observed metric and its null expectation divided by the standard deviation of the null distribution. These SES values were then used as alternative predictor variables in the species-level analyses, to evaluate the effects of phylogenetic distance on establishment after accounting for the influence of resident richness (*SI Appendix*, Fig. S9).

At the community level, successfully established species from all introduction events were classified as the established exotic community in each lake, while all resident species present prior to the first introduction were identified as the native community. For each lake, we calculated the mean pairwise phylogenetic distance (MPPD) and the mean nearest taxon distance (MNTD) between the established exotic and native communities (11, 16), providing complementary information on overall phylogenetic structure and nearest-neighbor relatedness. To assess the robustness of these community-level patterns, we additionally calculated SES. MPPD and SES.MNTD using two complementary null models (4, 15, 36). The first null model randomized the native community by sampling species from the full native pool while maintaining observed native richness and the exotic assemblage. The second randomized the exotic assemblage by sampling from the full exotic pool while holding the exotic richness and native community constant. SES values derived from these null models were then used as alternative response variables in subsequent community level analyses (*SI Appendix*, Fig. S16). To account for potential changes in native species composition over time, we additionally used native species observed in the final postintroduction period (i.e., the most recent records) as an alternative reference for the native community, and recalculated the phylogenetic distances between the exotic community and this updated native community. The results of this robustness analysis were consistent with those based on the initial native communities (Fig. 4 and *SI Appendix*, Fig. S15).

Estimation of Species Pool Size and Community Completeness. We estimated the dark diversity of each community using the hypergeometric co-occurrence-based method from the R package DarkDiv (39). Species pool size of each community was then measured as the sum of observed diversity and dark diversity. Community completeness was calculated as the ratio of observed diversity to dark diversity: ln (observed/dark diversity) (29, 31). Although both metrics are derived from observed and dark diversity, recent studies have demonstrated that these two metrics offer complementary insights and are often weakly correlated with each other (29, 58). The hypergeometric method infers the suitability of absent species for specific sites by comparing observed pairwise species co-occurrences with random expectations, thereby quantifying species associations under given community compositions. In this framework, consistently positive associations between an absent species and the species of the observed community indicate a higher likelihood that the species belongs to the site's dark diversity. Previous evaluations based on both simulated and empirical datasets indicate that the hypergeometric method provides well-calibrated estimates of species suitability, shows limited sensitivity to species frequency, and performs consistently across a wide range of sample sizes (38–40). In applications to empirical datasets, hypergeometric co-occurrence–based methods have been shown to capture broad patterns of species' environmental affinities and ecological associations, while avoiding extreme predictions and effectively distinguishing the suitability of common and rare species across sites (27, 39, 40).

By applying the hypergeometric method, we compared the observed co-occurrence of resident species across all communities with the random expectation described by the hypergeometric distribution (39, 69). The difference between observed and expected co-occurrences was standardized using the standard deviation of expected values, yielding effect sizes that quantify the strength of species pair associations (39, 40). For each absent resident species, we averaged its associations with all species present to estimate its probability of belonging to the community's dark diversity, ranging from 0 to 1 (38, 40). A probability close to 0.5 indicates neutral association, while values substantially above or below reflect stronger positive or negative deviations from random expectations (39). We calculated dark diversity using two thresholds for absent resident species: i) all predicted probabilities in the main text (27, 40), and ii) a stricter threshold (probability > 0.7), representing a more conservative species pool (39, 40), in *SI Appendix*. As a robustness analysis, we re-estimated dark diversity using the Beals-based favorability method. Specifically, we calculated Beals index for each absent native species based on its co-occurrence patterns with observed species across sites, transformed these indexes into favorability values to reduce the influence of regional species frequency, and summed favorability values across absent species to quantify dark diversity for each community [(38, 39) and *SI Appendix*, Figs. S6 and S14].

Because the Swedish lake dataset spans different bioregions, we further evaluated whether the definition of the regional species list influenced dark diversity estimates. Following the Infrastructure for Spatial Information in Europe (INSPIRE) Geoportal (https://inspire-geoportal.ec.europa.eu/), we assigned each lake to one of three bioregions: boreal, alpine, or continental. For each lake, the dark diversity was allowed to include only the species recorded in other lakes within the same bioregion. Variation in species availability among regions was further accounted for in the statistical models by including bioregion as a random effect. Results from this bioregion-restricted approach were highly consistent with those obtained using the full native species pool (Figs. 3 and 4 and *SI Appendix*, Figs. S5 and S12).

Statistical Analyses. At the species level, we assessed how the phylogenetic distance between exotic fish species and the native community affected their establishment success, and how this effect varied with site-specific species pool size and community completeness of native communities. We employed binomial generalized linear mixed models (GLMMs) with logit link functions to model establishment outcomes (success or failure) of introduced fishes, using the R package lme4 (70). The explanatory variables included MPD (or PNND), species pool size, and community completeness. We also tested for interaction effects between MPD (or PNND) and species pool size, as well as between MPD (or PNND) and community completeness. Because species pool size exhibited a pronounced negative skew, we applied a reflected-log-transformation prior to model fitting, in which values were reflected, log-transformed, and reflected back to preserve the original ordering. To account for the nonindependence of multiple introductions within each lake and repeated introductions of the same species, we incorporated lake and species identity as random effects. To further investigate whether lake area influences phylogenetic patterns of species establishment within the dark diversity framework, we expanded the GLMMs to include lake area as an additional explanatory variable. Specifically, we tested whether lake area directly affects species establishment and modifies the interaction between species pool size, community completeness, and exotic-native phylogenetic distance. To evaluate whether the dark diversity framework offers insights beyond species richness, we also reconstructed the GLMMs by substituting species pool size and community completeness with species richness, using species richness, MPD (or PNND), and their interactions as explanatory variables. These models were compared using the Akaike Information Criterion (AIC) and Akaike weights (AW), which estimate the relative likelihood that a given model is the best-supported model among the candidate set. We calculated standardized effect sizes for all explanatory variables and evaluated model fit using marginal R^2 (variation explained by fixed effects) and conditional R^2 (variation explained by both fixed and random effects) via the tab_model function in the sjPlot package (71, 72). Interaction effects were visualized using the visreg package (73), setting species pool size and community completeness values to the 10%, 30%, 50%, 70%, and 90% quantile across all lakes.

To address potential phylogenetic nonindependence among introduced species and ensure analytical robustness, we constructed a Bayesian phylogenetic hierarchical model. This model used the same response variable, explanatory variables, and random effects as the GLMMs, but additionally incorporated a phylogenetic covariance structure. We generated the phylogenetic covariance matrix for the entire phylogeny using the R package ape and extracted the submatrix specific to the introduced species. The model was built using the R package INLA (74), which employs integrated nested Laplace approximation for Bayesian inference. We applied default INLA priors for all fixed effect parameters and log-gamma priors for the random effect parameters. Results from the Bayesian phylogenetic hierarchical models were consistent with those from the GLMMs, reinforcing the reliability of our findings. Therefore, we present the results of the GLMMs in the main text, while the Bayesian model results are provided in *SI Appendix*, Tables S2 and S4.

At the community level, we examined how phylogenetic distances between all successfully established exotic species and the native species in each lake were influenced by native species pool size, community completeness, and additional lake characteristics. Using the R package lme4, we constructed linear models with the MPPD (or MNTD) between all established exotic and native species in each lake as the response variable. We used species pool size and community completeness as primary explanatory variables, and included lake area, latitude, elevation, and accumulated water temperature to assess whether the dark diversity framework more effectively explains the context dependence of exotic-native phylogenetic distances. Both MPPD and MNTD models showed approximate normality, with only minor deviations apparent in quantile–quantile plots (*SI Appendix*, Fig. S17). Similarly, we calculated standardized effect sizes for all explanatory variables via the tab_model function (71).

At both species and community levels, we found that the results based on nearest distance indices (PNND and MNTD) were largely consistent with those from mean distance indices (MPD and MPPD). Therefore, we present only the mean distance indexes in the main text, with nearest distance indexes included in *SI Appendix*, Figs. S2 and S10 and Tables S3 and S4. All statistical analyses were conducted in R v.4.2.3 (75).

Data, Materials, and Software Availability. The data and code used in this study are available in Figshare (https://doi.org/10.6084/m9.figshare.28369766) (76). The Swedish fish introduction records are sourced from the data paper published by Henriksson et al. (47). We obtained molecular phylogenetic data for all fish species from Xu et al. (63).

ACKNOWLEDGMENTS. We thank Göran Englund for his helpful clarifications regarding data-related questions and his valuable suggestions. We also appreciate the insightful feedback from Xingfeng Si, Marc W. Cadotte, and Mark van Kleunen, which significantly contributed to improving the quality of this manuscript. We are also grateful to the anonymous reviewers for their constructive comments, which substantially enhanced this study. This research was supported by the National Key R&D Program (2023YFC2604500), the National Natural Science Foundation of China (NSFC32571947, 32222051, and 31971553), the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (JYB2025XDXM904), ECNU Multifunctional Platform for Innovation (001). M.P. was funded by the Estonian Research Council (PRG3127), and the Ministry of Education and Research of Estonia (Center of Excellence AgroCropFuture, TK200). S.Y.F. was supported by the project Center for Biological Invasions in Forests (101187384), funded by the European Union through the Horizon Europe programme.

Author affiliations: [a]Research Center for Global Change and Ecological Forecasting, Zhejiang Tiantong Forest Ecosystem National Observation and Research Station, Institute of Eco-Chongming, School of Ecological and Environmental Sciences, East China Normal University, Shanghai 200241, China; [b]Key Laboratory of Lake and Watershed Science for Water Security, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China; [c]Key Laboratory of Prevention and Control for Aquatic Invasive Alien Species, Ministry of Agriculture and Rural Affairs, Pearl River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Guangzhou 510380, China; [d]Ecology, Department of Biology, University of Konstanz, Konstanz 78457, Germany; [e]The Forest Risk Research Centre, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Prague 16500, Czech Republic; [f]Department of Plant Sciences, University of Saskatchewan, Saskatoon, SK S7N5A8, Canada; [g]Research Centre in Sustainable Energy, Institute for Global Food Security, School of Biological Sciences, Queen's University Belfast, Belfast BT95DL, United Kingdom; and [h]Department of Botany, Institute of Ecology and Earth Science, University of Tartu, Tartu 50409, Estonia

1. P. Pyšek *et al.*, Scientists' warning on invasive alien species. *Biol. Rev.* **95**, 1511–1534 (2020).
2. T. Peller, F. Altermatt, Invasive species drive cross-ecosystem effects worldwide. *Nat. Ecol. Evol.* **8**, 1087–1097 (2024).
3. H. E. Roy *et al.*, Curbing the major and growing threats from invasive alien species is urgent and achievable. *Nat. Ecol. Evol.* **8**, 1216–1223 (2024).
4. L. Gallien, M. Carboni, The community ecology of invasive species: Where are we and what's next? *Ecography* **40**, 335–352 (2017).
5. G. Su, A. Mertel, S. Brosse, J. M. Calabrese, Species invasiveness and community invasibility of North American freshwater fish fauna revealed via trait-based analysis. *Nat. Commun.* **14**, 2332 (2023).
6. D. M. Richardson, P. Pysek, Naturalization of introduced plants: Ecological drivers of biogeographical patterns. *New Phytol.* **196**, 383–396 (2012).
7. E. Z. Daly *et al.*, A synthesis of biological invasion hypotheses associated with the introduction–naturalisation–invasion continuum. *Oikos* **2023**, e09645 (2023).
8. J. A. Bennett, M. Pärtel, Predicting species establishment using absent species and functional neighborhoods. *Ecol. Evol.* **7**, 2223–2237 (2017).
9. J. L. Lockwood, P. Cassey, T. Blackburn, The role of propagule pressure in explaining species invasions. *Trends Ecol. Evol.* **20**, 223–228 (2005).
10. C. Darwin, *On the Origin of Species* (J. Murray, London, 1859).
11. M. W. Cadotte, S. E. Campbell, S. P. Li, D. S. Sodhi, N. E. Mandrak, Preadaptation and naturalization of nonnative species: Darwin's two fundamental insights into species invasion. *Annu. Rev. Plant Biol.* **69**, 661–684 (2018).
12. A. Ricciardi, M. Mottiar, Does Darwin's naturalization hypothesis explain fish invasions? *Biol. Invasions* **8**, 1403–1407 (2006).
13. B. S. Maitner, J. A. Rudgers, A. E. Dunham, K. D. Whitney, Patterns of bird invasion are consistent with environmental filtering. *Ecography* **35**, 614–623 (2012).
14. C. C. Daehler, Darwin's naturalization hypothesis revisited. *Am. Nat.* **158**, 324–330 (2001).
15. J. M. Diez, J. J. Sullivan, P. E. Hulme, G. Edwards, R. P. Duncan, Darwin's naturalization conundrum: Dissecting taxonomic patterns of species invasions. *Ecol. Lett.* **11**, 674–681 (2008).
16. W. Thuiller *et al.*, Resolving Darwin's naturalization conundrum: A quest for evidence. *Divers. Distrib.* **16**, 461–475 (2010).
17. J. N. Pinto-Ledezma *et al.*, Testing Darwin's naturalization conundrum based on taxonomic, phylogenetic, and functional dimensions of vascular plants. *Ecol. Monogr.* **90**, e01420 (2020).
18. C. Ma *et al.*, Different effects of invader-native phylogenetic relatedness on invasion success and impact: A meta-analysis of Darwin's naturalization hypothesis. *Proc. R. Soc. B. Biol. Sci.* **283**, 20160663 (2016).
19. J. Wang *et al.*, Darwin's naturalization conundrum reconciled by changes of species interactions. *Ecology* **104**, e3850 (2022).
20. B. S. Rocha, M. V. Cianciaruso, Water temperature and lake size explain Darwin's conundrum for fish establishment in boreal lakes. *Hydrobiologia* **848**, 2033–2042 (2021).
21. S. Y. Fan *et al.*, A latitudinal gradient in Darwin's naturalization conundrum at the global scale for flowering plants. *Nat. Commun.* **14**, 6244 (2023).
22. D. S. Park, X. Feng, B. S. Maitner, K. C. Ernst, B. J. Enquist, Darwin's naturalization conundrum can be explained by spatial scale. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 10904–10910 (2020).
23. Afshana, J. N. Pinto-Ledezma, Z. A. Reshi, phylogenetic relatedness of plant species co-occurring with an invasive alien plant species varies with elevation. *Ecosphere* **15**, e4966 (2024).
24. G. Wang *et al.*, The impact of species phylogenetic relatedness on invasion varies distinctly along resource versus non-resource environmental gradients. *J. Appl. Ecol.* **61**, 869–883 (2024).
25. Y. Feng, T. D. Fouqueray, M. van Kleunen, Linking Darwin's naturalisation hypothesis and Elton's diversity–invasibility hypothesis in experimental grassland communities. *J. Ecol.* **107**, 794–805 (2019).
26. K. Mokany, D. R. Paini, Dark diversity: Adding the grey. *Trends Ecol. Evol.* **26**, 264–265 (2011).
27. M. Pärtel *et al.*, Global impoverishment of natural vegetation revealed by dark diversity. *Nature* **641**, 917–924 (2025).
28. M. Zobel, The species pool concept as a framework for studying patterns of plant diversity. *J. Veg. Sci.* **27**, 8–18 (2016).
29. M. Pärtel, R. Szava-Kovats, M. Zobel, Community completeness: Linking local and dark diversity within the species pool concept. *Folia Geobot.* **48**, 307–317 (2013).
30. C. Flojgaard *et al.*, Dark diversity reveals importance of biotic resources and competition for plant diversity across habitats. *Ecol. Evol.* **10**, 6078–6088 (2020).
31. K. Riibak *et al.*, Drivers of plant community completeness differ at regional and landscape scales. *Agric. Ecosyst. Environ.* **301**, 107004 (2020).
32. C. Elton, *The Ecology of Invasions by Animals and Plants* (Methuen, London, 1958).
33. J. J. Wiens, C. H. Graham, Niche conservatism: Integrating evolution, ecology, and conservation biology. *Annu. Rev. Ecol. Evol. Syst.* **36**, 519–539 (2005).
34. P. Gauzere *et al.*, Vacant yet invasible niches in forest community assembly. *Funct. Ecol.* **34**, 1945–1955 (2020).
35. C. Bernery *et al.*, Freshwater fish invasions: A comprehensive review. *Annu. Rev. Ecol. Evol. Syst.* **53**, 427–456 (2022).
36. D. S. Park, D. Potter, A test of Darwin's naturalization hypothesis in the thistle tribe shows that close relatives make bad neighbors. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17915–17920 (2013).
37. S. P. Li *et al.*, The effects of phylogenetic relatedness on invasion success and impact: Deconstructing Darwin's naturalisation conundrum. *Ecol. Lett.* **18**, 1285–1292 (2015).
38. R. J. Lewis, R. Szava-Kovats, M. Pärtel, Estimating dark diversity and species pools: An empirical assessment of two methods. *Methods Ecol. Evol.* **7**, 104–113 (2016).
39. C. P. Carmona, M. Pärtel, Estimating probabilistic site-specific species pools and dark diversity from co-occurrence data. *Glob. Ecol. Biogeogr.* **30**, 316–326 (2021).
40. B. Paganeli, J. Fujinuma, D. P. F. Trindade, C. P. Carmona, M. Pärtel, A roadmap to carefully select methods for dark-diversity studies. *J. Veg. Sci.* **35**, e13264 (2024).
41. R. E. Gozlan, J. R. Britton, I. Cowx, G. H. Copp, Current knowledge on non-native freshwater fish introductions. *J. Fish Biol.* **76**, 751–786 (2010).
42. J. Cucherousset, J. D. Olden, Ecological impacts of nonnative freshwater fishes. *Fisheries* **36**, 215–230 (2011).
43. M. Xu *et al.*, Global freshwater fish invasion linked to the presence of closely related species. *Nat. Commun.* **15**, 1411 (2024).
44. P. B. Moyle, M. P. Marchetti, Predicting invasion success: Freshwater fishes in California as a model. *BioScience* **56**, 515–524 (2006).
45. A. C. Rodrigues, J. Cucherousset, E. R. Cunha, N. C. L. dos Santos, L. C. Gomes, Functional dissimilarity correlates to the co-occurrence patterns of native and non-native species. *Biol. Invasions.* **26**, 2475–2487 (2024).
46. G. Su, S. Villéger, S. Brosse, Morphological sorting of introduced freshwater fish species within and between donor realms. *Glob. Ecol. Biogeogr.* **29**, 803–813 (2020).
47. A. Henriksson, C. Rydberg, G. Englund, Failed and successful intentional introductions of fish species into 821 Swedish lakes. *Ecology* **97**, 1364–1364 (2016).
48. H. V. Cornell, S. P. Harrison, What are species pools and when are they important? *Annu. Rev. Ecol. Evol. Syst.* **45**, 45–67 (2014).

49. J. A. Catford *et al.*, The intermediate disturbance hypothesis and plant invasions: Implications for species richness and management. *Perspect. Plant Ecol. Evol. Syst.* **14**, 231–241 (2012).
50. A. Ronk, R. Szava-Kovats, M. Zobel, M. Pärtel, Observed and dark diversity of alien plant species in Europe: Estimating future invasion risk. *Biodivers. Conserv.* **26**, 899–916 (2017).
51. I. M. Parker *et al.*, Phylogenetic structure and host abundance drive disease pressure in communities. *Nature* **520**, 542–544 (2015).
52. D. B. Fitzgerald, M. Tobler, K. O. Winemiller, From richer to poorer: Successful invasion by freshwater fishes depends on species richness of donor and recipient basins. *Glob. Chang. Biol.* **22**, 2440–2450 (2016).
53. C. C. Muhlfeld *et al.*, Legacy introductions and climatic variation explain spatiotemporal patterns of invasive hybridization in a native trout. *Glob. Chang. Biol.* **23**, 4663–4674 (2017).
54. G. A. Polis, C. A. Myers, R. D. Holt, The ecology and evolution of intraguild predation: Potential competitors that eat each other. *Annu. Rev. Ecol. Evol. Syst.* **20**, 297–330 (1989).
55. D. A. Jackson, P. R. Peres-Neto, J. D. Olden, What controls who is where in freshwater fish communities—The roles of biotic, abiotic, and spatial factors. *Can. J. Fish. Aquat. Sci.* **58**, 157–170 (2001).
56. G. V. Braoudakis, D. A. Jackson, Effect of lake size, isolation and top predator presence on nested fish community structure. *J. Biogeogr.* **43**, 1425–1435 (2016).
57. T. Mehner, S. Brucet, "Structure of fish communities in lakes and its abiotic and biotic determinants" in *Encyclopedia of Inland Waters*, T. Mehner, K. Tockner, Eds. (Elsevier, Oxford, 2022), pp. 77–88.
58. J. A. Bennett *et al.*, Species pools, community completeness and invasion: Disentangling diversity effects on the establishment of native and alien species. *Ecol. Lett.* **19**, 1496–1505 (2016).
59. D. Xing, J. Zhang, F. He, Comment on "interspecific competition limits bird species' ranges in tropical mountains". *Science* **379**, eade2109 (2023).
60. R. E. Ricklefs, Community diversity: Relative roles of local and regional processes. *Science* **235**, 167–171 (1987).
61. S. P. Li, J. Tan, X. Yang, C. Ma, L. Jiang, Niche and fitness differences determine invasion success and impact in laboratory bacterial communities. *ISME J.* **13**, 402–412 (2019).
62. J. Divíšek *et al.*, Similarity of introduced plant species to native ones facilitates naturalization, but differences enhance invasion success. *Nat. Commun.* **9**, 4631 (2018).
63. M. Xu *et al.*, Exotic fishes that are phylogenetically close but functionally distant to native fishes are more likely to establish. *Glob. Chang. Biol.* **28**, 5683–5694 (2022).
64. A. Henriksson, D. A. Wardle, J. Trygg, S. Diehl, G. Englund, Strong invaders are strong defenders—Implications for the resistance of invaded communities. *Ecol. Lett.* **19**, 487–494 (2016).
65. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
66. S. W. Kembel *et al.*, Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463–1464 (2010).
67. H. Schaefer, O. J. Hardy, L. Silva, T. G. Barraclough, V. Savolainen, Testing Darwin's naturalization hypothesis in the Azores. *Ecol. Lett.* **14**, 389–396 (2011).
68. S. Y. Strauss, C. O. Webb, N. Salamin, Exotic taxa less related to native species are more invasive. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5841–5845 (2006).
69. H. T. Arita, Species co-occurrence analysis: Pairwise versus matrix-level approaches. *Glob. Ecol. Biogeogr.* **25**, 1397–1400 (2016).
70. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
71. D. Lüdecke, sjPlot–Data visualization for statistics in social science. *R. package version* **2.8.16**, 1–107 (2018).
72. S. Nakagawa, H. Schielzeth, A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods Ecol. Evol.* **4**, 133–142 (2013).
73. P. Breheny, W. Burchett, Visualization of regression models using visreg. *R J.* **9**, 56 (2017).
74. H. Rue, S. Martino, N. Chopin, Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. B.* **71**, 319–392 (2009).
75. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing Vienna, Austria, 2024).
76. W. Zhang *et al.*, Data and R codes for Zhang et al. "Dark diversity framework reconciles Darwin's naturalization conundrum for freshwater fish invasions". Figshare. https://doi.org/10.6084/m9.figshare.28369766. Deposited 30 September 2025.