


A unified framework to investigate and interpret hybrid and allopolyploid biodiversity across biological scales

Christopher P. Krieg 

Department of Botany, University of Wisconsin, Madison, Wisconsin, USA

Correspondence

Christopher P. Krieg, Department of Botany, University of Wisconsin, Madison, Wisconsin, USA.

Email: christopher.p.krieg@gmail.com

This article is part of the special issue “Twice as Nice: New Techniques and Discoveries in Polyploid Biology.”

Abstract

Premise: Hybridization and polyploidization are common in vascular plants and important drivers of biodiversity by facilitating speciation and ecological diversification. A primary limitation to making broad synthetic discoveries in hybrid and allopolyploid biodiversity research is the absence of a standardized framework to compare data across studies and biological scales.

Methods: Here, I present a new quantitative framework to investigate and interpret patterns in hybrid and allopolyploid biology called the divergence index (DI). The DI framework produces standardized data that are comparable across studies and variables. To show how the DI framework can be used to synthesize data, I analyzed published biochemical, physiological, and ecological trait data of hybrids and allopolyploids. I also apply key ecological and evolutionary concepts in hybrid and polyploid biology to translate nominal outcomes, including transgression, intermediacy, expansion, and contraction, in continuous DI space.

Results: Biochemical, physiological, ecological, and evolutionary data can all be analyzed, visualized, and interpreted in the DI framework. The DI framework is particularly suited to standardize and compare variables with very different scales. When using the DI framework to understand niche divergence, a metric of niche overlap can be used to complement insights to centroid and breadth changes.

Discussion: The DI framework is an accessible framework for hybrid and allopolyploid biology and represents a flexible and intuitive tool that can be used to reconcile outstanding problems in plant biodiversity research.

KEYWORDS

biodiversity, divergence index, ecology, hybrids, niche, polyploids

Hybridization and polyploidization have impacted the evolution of most major plant groups and are thought to be important drivers of plant biodiversity including as a mechanism for speciation (Winge, 1917; Grant, 1981; Soltis and Soltis, 2009; Jiao et al., 2011; Mayrose et al., 2011; Nieto Feliner et al., 2020), physiological novelty (Maherali et al., 2009; Hao et al., 2013; Manzaneda et al., 2015; Corneille et al., 2019), and ecological diversification (Stebbins, 1985; Hijmans et al., 2007; Ramsey, 2011; Baniaga et al., 2020; Singhal et al., 2021). Despite the importance of hybridization and polyploidization to biodiversity, some of the most basic aspects of their biology remain poorly understood. For

example, making connections between the mechanisms that underlie hybrid and allopolyploid biological outcomes from genes to ecology has been especially challenging. A primary lens through which we understand polyploidization as a driver of species diversification is through the immediate introduction of reproductive barriers and new combinations and/or functions of genetic material (Levin, 1983; Thompson and Lumaret, 1992; Otto and Whitton, 2000). New combinations of genetic material are thought to lead to new phenotypes, which can impact species ecology and thus generate spatial plant biodiversity (Bombliès, 2020; Van de Peer et al., 2021). Some studies have shown that hybridization

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2024 The Author(s). *Applications in Plant Sciences* published by Wiley Periodicals LLC on behalf of Botanical Society of America.

between more phylogenetically and ecologically distinct taxa can cause greater genetic disruption in hybrids compared to those formed between more similar progenitor taxa (Buggs et al., 2009; Paun et al., 2009; Stelkens and Seehausen, 2009). However, the differentiation between progenitors is rarely explicitly considered in empirical studies of allopolyploid biology. Consider the hypothetical example of two different sets of progenitors on terrestrial Earth; one set of progenitors, D_1 and D_2 , occur close together in environmental niche space, while another set of progenitors, D_3 and D_4 , are more differentiated in their niches (Figure 1A). If ecological divergence is associated with genetic divergence within each set of progenitors, and both sets of progenitors form an allopolyploid, what is the relative potential for new combinations and/or functions of genetic material in their polyploids? Consistent with the few existing empirical studies (Clausen et al., 1945; Stelkens and Seehausen, 2009), it seems reasonable to posit that progenitors with more divergent genomes may be more likely to produce hybrid polyploids with novel combinations of genetic material, relative to progenitors that are genetically more similar to each other. Conversely, if the two sets of hypothetical progenitors (one similar set D_1 and D_2 , and one very different set D_3 and D_4) each formed an allopolyploid with the same characteristics, our perspective on the relative impact of polyploidization on plant ecology and evolution would be very different between these cases (Figure 1B). This suggests that our tools to understand the mechanisms that drive variation in polyploid ecological outcomes should be explicitly standardized by the differences between progenitors in each hybrid and/or allopolyploid complex.

The concept of standardizing interpretations of polyploid characteristics by the progenitor characteristics has roots in polyploid biology theory. A key tenet of polyploid biology theory is that the relationship (e.g., dis/similarity) between polyploid and progenitor characteristics (e.g., niche occupancy, competitive ability, stress tolerance, reproductive

mode) strongly determines the ability of newly formed polyploids to establish and persist in the wild (Stebbins, 1971; Levin, 1975; Soltis et al., 2014). While previous research has made some progress in understanding the physiological mechanisms that drive polyploid niche separation from progenitors within a single complex (Hao et al., 2013; Zhang et al., 2017; Losada et al., 2023), unifying environmental, ecological, and physiological data to characterize the mechanisms that underlie patterns of polyploid biodiversity has been challenging (Soltis et al., 2010, 2014; Parisod and Broennimann, 2016; Gaynor et al., 2020; Shimizu, 2022). Ultimately, if we hope to better characterize the mechanisms that drive variation in polyploid biology and the role of polyploidization in generating plant biodiversity, our methods require a stronger synthesis and unification of data across biological scales.

A common practice in biology is the assignment of hybrid and allopolyploid biological outcomes to nominal categories such as intermediacy (between progenitors), expansion and contraction (an increase or decrease in variation), and transgression (very different from progenitors) (Brochmann et al., 2004; Hijmans et al., 2007; McIntyre, 2012; Glennon et al., 2014; Harbert et al., 2014; Marchant et al., 2016). While there is strong consensus on the conceptual nature of these categories (i.e., intermediacy, expansion, contraction, and transgression) (Figure 2), several different qualitative and quantitative thresholds have been used to classify these key biological outcomes (Anderson, 1953; Szymura and Barton, 1991; Rieseberg et al., 1998; Anderson and Thompson, 2002; Thompson et al., 2021), limiting our ability to compare results across studies that use different definitions and data. In studies of hybrid and polyploid ecology, the relative contributions of changes in niche breadth and centroid position to the categorization of a polyploid's niche occupancy are often confounded and can be difficult to untangle (see Glennon et al., 2014). For example, a large shift in niche centroid position (i.e., central

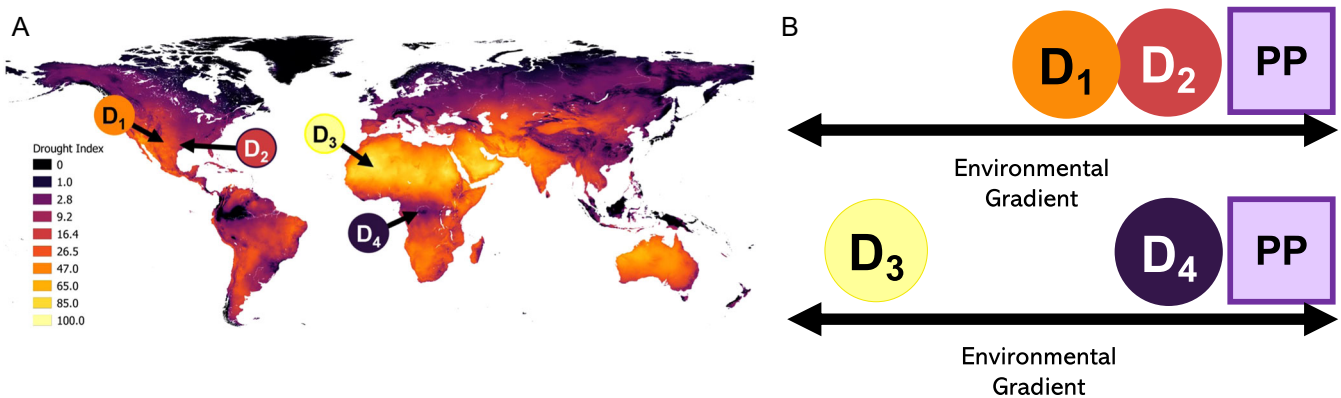


FIGURE 1 Two hypothetical cases demonstrating the impact of progenitor similarity on our interpretations of the evolutionary and ecological significance of polyploidy. (A) Consider the progenitors D_1 and D_2 with similar drought indices and the more dissimilar progenitors D_3 and D_4 , where each set of progenitors forms an allopolyploid (PP) with the same characteristics. (B) The allopolyploid complex with D_1 and D_2 shows greater ecological and potential evolutionary change compared to the polyploid complex with D_3 and D_4 and thus demonstrates that the difference between progenitors can impact our view of polyploidization as a driver of biodiversity. The global drought index raster is from McCulloh et al. (2023) and represents environments with very low water availability (100) to very high water availability (0).

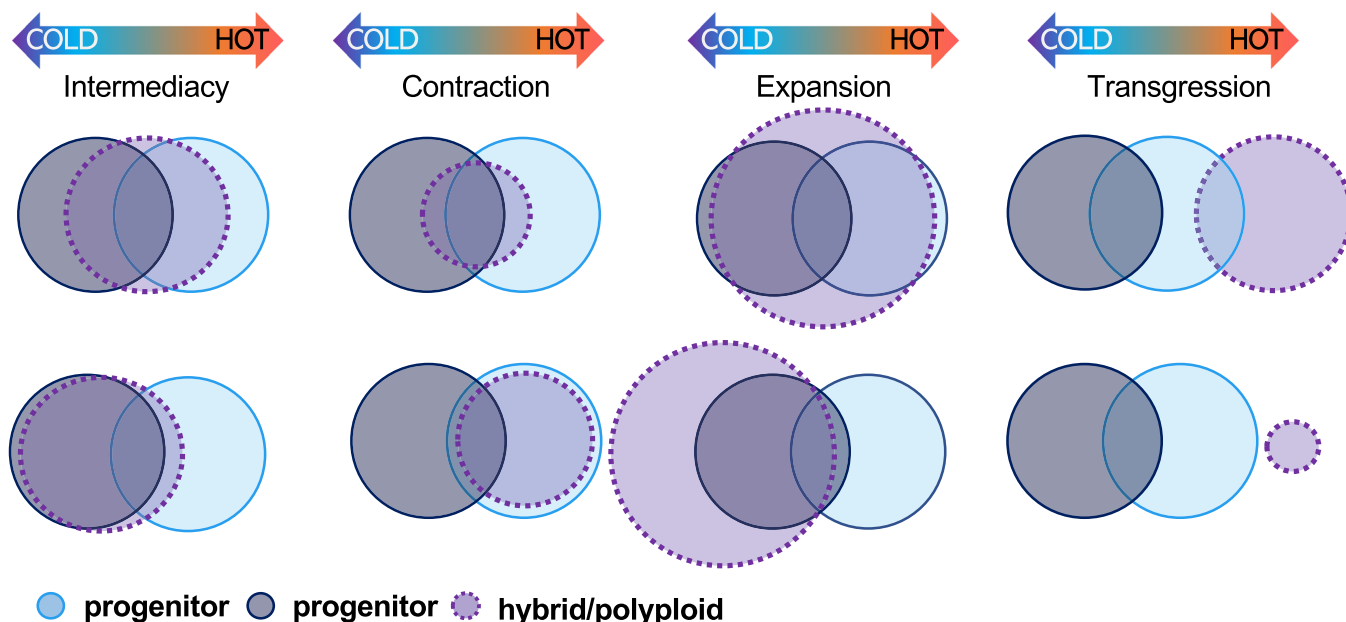


FIGURE 2 Conceptual diagram of key ecological outcomes thought to be important in hybrid and polyploid ecology and evolution. The top row shows a simplified consensus depiction of intermediacy, contraction, expansion, and transgression along a hypothetical niche axis of temperature variation. The bottom row shows plausible changes in niche centroid and breadth that make classifications more obscure, more subjective, and less comparable. Light blue and gray circles represent progenitors, and purple circles with dotted borders represent hybrid and/or allopolyploid offspring.

tendency) beyond progenitors, with or without a change in niche breadth (i.e., variation), can result in a polyploid niche that is highly divergent from progenitors (i.e., apparent transgression). When a polyploid has a highly divergent niche centroid and a highly expanded niche breadth, should it be classified as niche transgression, expansion, or both? If changes in niche breadth and in niche centroid position can independently or concurrently produce the ecological patterns observed for polyploids, how can we clearly and quantitatively distinguish between multiple potential mechanisms that drive polyploid niche outcomes? Clearly capturing changes in niche centroid position independently of changes in niche breadth in continuous variable space, as well as their contribution to categorical classifications, is critical to disentangling the underlying drivers of hybrid and allopolyploid ecological patterns. A similar clarity is necessary to understand the drivers of hybrid and allopolyploid anatomical, physiological, and ecological outcomes that determine spatial patterns of biodiversity.

Here, I formalize a quantitative framework, called the divergence index (hereafter “DI”), to unify hybrid and allopolyploid biology data and permit direct comparisons across datasets and measurement types. In essence, the DI is a measure of a hybrid or allopolyploid’s characteristics as a function of the difference between the progenitors’ characteristics (Eq. 1; Figure 3). In its analytical form, the DI framework is a two-dimensional space that represents a holistic system to quantify, visualize, and interpret changes in hybrid and/or allopolyploid central position (e.g., centroid) and variation (e.g., breadth). To demonstrate, I apply the DI framework to published data that differ in type and scale,

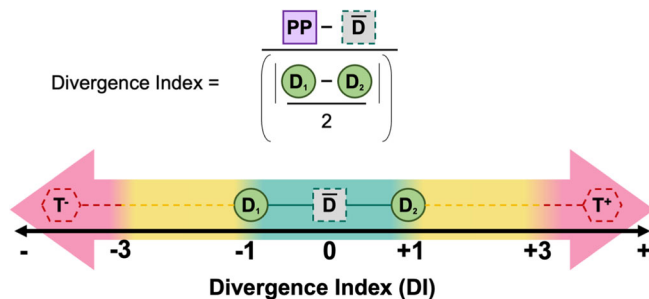


FIGURE 3 The generic formulation of the divergence index (DI) concept and the resulting DI scale, where PP is the polyploid’s value (which can be any continuous data type), \bar{D} is the arithmetic mean of progenitor values, and D_1 and D_2 are the two progenitor species. This formulation results in an indexed scale where 0 represents a polyploid that is perfectly intermediate between the progenitor values. DI values of -1 and 1 represent the lower and the higher progenitor values, respectively. DI values of -3 and 3 represent polyploid differentiation beyond progenitor values equal to the difference between the progenitors themselves. Note that this figure is meant to convey the conceptual function of the DI framework, and the use of the two-dimensional DI framework is recommended.

including biochemical, physiological, morphological, ecological, and environmental data, and reveal or recapitulate important insights to hybrid and allopolyploid biology. I also translate key concepts in polyploid biology including intermediacy, expansion, contraction, transgression, and niche shifts within the DI framework and use environmental data from extant allopolyploid complexes as examples. I discuss the rationale and mathematical approach behind the translation of these classifications into continuous variable space

(comprised of DI axes; see below) and their strengths and current limitations.

A simple framework for hybrid and allopolyploid biology

Most of the tools currently used to compare inter-specific hybrids and allopolyploids were designed for pairwise comparisons between two entities. When used in a three-entity complex, this leads to three or more metrics of differentiation that must be calculated and interpreted (e.g., a polyploid vs. progenitor 1, then polyploid vs. progenitor 2, then progenitor 1 vs. progenitor 2). In its simplest conceptual form, the DI reduces this complexity to a single measure of a hybrid or allopolyploid's characteristics relative to the difference between progenitor characteristics:

$$DI = \frac{PP - \bar{D}}{\left| \frac{D_1 - D_2}{2} \right|} \quad (1)$$

where PP is the polyploid, \bar{D} represents “perfect” intermediacy and is the arithmetic mean between the progenitor input values, D_1 is one of the progenitor species, and D_2 is the other progenitor species. This conceptual formulation results in a standardized scale where $DI = 0$ represents a hybrid that is perfectly intermediate between the progenitor input values (Figure 3). DI values of -1 and 1 represent a PP equal to the lower and the higher progenitor input values, respectively. The DI values of -3 and 3 represent the divergence beyond the progenitor values where a hybrid is as different from its progenitors as the progenitor species are from each other. With this formulation, one unit in DI space is equal to one-half the difference between progenitor values.

To characterize the biological divergence of a plant hybrid and/or allopolyploid from its progenitors, the two-dimensional analytical form of the DI framework uses two key aspects of data distributions—central tendency (i.e., centroid) and variation (i.e., breadth)—as inputs. In the DI framework, centroid divergence (DI_C) is quantified as:

$$DI_C = \frac{PP_C - \bar{D}_C}{\left| \frac{D_{1,C} - D_{2,C}}{2} \right|} \quad (2)$$

where the subscript “C” denotes the output as centroid divergence (DI_C) or input as the centroid of the data distribution (e.g., mean or median). PP is the hybrid or allopolyploid, \bar{D} is perfect intermediacy between progenitors (mean of progenitor centroids), and D_1 and D_2 are the progenitor centroids. Similarly, breadth divergence (DI_B) is quantified as:

$$DI_B = \frac{PP_B - \bar{D}_B}{\left| \frac{D_{1,B} - D_{2,B}}{2} \right|} \quad (3)$$

where the subscript “B” denotes the output as breadth divergence (DI_B) or input as the breadth of the data distribution (e.g., standard deviation). PP is the hybrid or allopolyploid, \bar{D} is perfect intermediacy between progenitors, and D_1 and D_2 are the progenitor breadths.

The centroid and breadth input values to the DI framework can be from any continuous data type (e.g., genetic similarity, morphology, metabolite production, photosynthetic capacity, plant height, climate niche, etc.) (see Video S1).

Translating key concepts in hybrid and allopolyploid biology

By using measures of variation and central tendency (breadth: DI_B , and centroid position: DI_C), I translated to DI space five of the most common conceptual biological patterns thought to impact polyploid biodiversity patterns and their evolutionary legacy, including (1) transgression, (2) intermediacy, (3) contraction, (4) expansion, and (5) niche shifts (or trait shifts, more generally).

In the DI framework, the threshold for an allopolyploid that has diverged beyond progenitor values occurs at a $DI_{C,B}$ less than -1 or greater than 1 , while the threshold for an allopolyploid that has diverged beyond progenitor values by a difference equal to the difference between its progenitor taxa always occurs at a $DI_{C,B}$ of -3 or 3 (see Figure 3). I use this stable benchmark of -3 or 3 as a strict threshold to further discriminate the magnitude of biological patterns (see the sections below on transgression, contraction, expansion, and niche shift). Specifically, I assign the prefix *hyper-* (from Greek “hyper,” meaning over, beyond, excess) to allopolyploid taxa that exhibit strong divergence greater than the divergence between progenitor taxa. Conversely, in the DI framework, an allopolyploid with characteristics beyond progenitor values but not more different than the progenitors are from each other is denoted with the prefix *hypo-* (from Greek “hypo,” meaning under, beneath, less than). These additional zones aid in recognizing meaningful changes in allopolyploid biological outcomes that do not exceed the relatively strict thresholds of the main zones (i.e., $DI_{B,C} \pm 3$ in definitions of transgression, contraction, expansion, and niche shift). Below, I focus on the application of these concepts in an ecological context and use ecological language and examples; however, the classifications below can be made with any continuous data as inputs and need not be ecological data.

Transgression: Transgression is a widely used concept in hybrid and polyploid biology to describe characteristics (e.g., gene expression, phenotypes, ecology) of a hybrid and/or polyploid that are highly divergent from the progenitor(s) from which it originated. In polyploid biology theory, transgression (and, similarly, trait shifts) is thought to be among the most critical factors that impact the survival, persistence, and ecological success of newly formed polyploids (e.g., Levin, 1975; Fowler and Levin, 1984, 2016).

The specific quantitative thresholds used to classify a polyploid's biological characteristics relative to the progenitor species vary across studies. For example, some previous research on polyploid niche divergence has set the threshold for allopolyploid transgression as an allopolyploid niche that is more different from its progenitors than the average progenitor difference in a dataset of multiple allopolyploid complexes (e.g., Marchant et al., 2016). Many previous studies have made various use of the difference between progenitor values to understand hybrid and/or polyploid biology (Marchant et al., 2016; Baniaga et al., 2020; Wang et al., 2022) (see Figure 1). Here, I adapt this concept to apply to each hybrid complex individually in the DI framework with respect to changes in centroids and breadths (see below for a discussion of overlap). This approach maintains the conceptual reasoning for comparing polyploids relative to progenitor differences (i.e., roots in polyploid biology theory), but avoids the potential for a polyploid characteristic to be classified one way (e.g., as transgressive) when part of a particular dataset but classified differently (e.g., not transgressive) if included in a different dataset when the mean difference between progenitors differs between datasets (see Marchant et al., 2016).

In hybrid and/or polyploid ecology, it is often difficult to separate multiple changes in niche occupancy that may be occurring simultaneously. For example, an allopolyploid may exhibit such strong expansion that both its breadth and centroid position are far beyond progenitor values, yet this scenario is often simply and subjectively referred to as “transgression” or as a “niche shift.” Similarly, a polyploid with a strongly contracted niche breadth may be far beyond progenitor niche limits, thereby presenting the question: Should the polyploid's ecology be classified as transgression, niche contraction, a niche shift, or some combination? One key benefit to the DI framework is that changes in niche centroid and breadth are clearly distinguishable. In other words, the DI framework directly informs the relative contribution of changes in niche breadth and centroid to hybrid and allopolyploid niche divergence from progenitors. Specifically, the DI framework recognizes all possible combinations of niche centroid (i.e., DI_C) and breadth (i.e., DI_B) differences between polyploids and their progenitors, thereby revealing not one but four main zones (i.e., types) of hyper-transgression (Figure 4). I arbitrarily ordered the four DI hyper-transgression zones (hyper-T1–4) following the general spatial arrangement of the four quadrants in a two-dimensional Cartesian plane:

hyper-transgression zone 1 (hyper-T1): A polyploid with a DI niche centroid (DI_C) and DI niche breadth (DI_B) that are both greater than 3 such that:

$$DI_C > 3 \text{ and } DI_B > 3 \quad (4)$$

hyper-transgression zone 2 (hyper-T2): A polyploid with a DI niche centroid (DI_C) that is less than -3 and DI niche breadth (DI_B) that is greater than 3 such that:

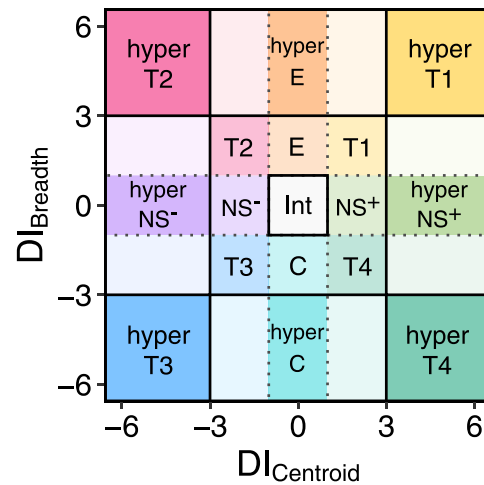


FIGURE 4 Four traditional ecological outcomes thought to be important in polyploid ecology and evolution and 21 additional outcomes revealed using the divergence index (DI). Intermediacy (Int) is in the center gray region, contractions (C) in the bottom center teal regions, expansions (E) in the upper orange regions, and niche shifts (NS) in negative (NS⁻) left purple regions and positive (NS⁺) right green regions. The eight main types of transgression (hypo- and hyper-T1–4) are in the corner regions. Eight additional outer regions (those in pale colors) represent outcomes with $DI_{C,B}$ values beyond ± 3 on one axis but not both. By convention, DI_C is placed on the x -axis.

$$DI_C < -3 \text{ and } DI_B > 3 \quad (5)$$

hyper-transgression zone 3 (hyper-T3): A polyploid with a DI niche centroid (DI_C) that is less than -3 and DI niche breadth (DI_B) that is less than -3 such that:

$$DI_C < -3 \text{ and } DI_B < -3 \quad (6)$$

hyper-transgression zone 4 (hyper-T4): A polyploid with a DI niche centroid (DI_C) that is greater than 3 and DI niche breadth (DI_B) that is less than -3 such that:

$$DI_C > 3 \text{ and } DI_B < -3 \quad (7)$$

The four zones of hyper-transgression (hyper-T1–4) arise from the added resolution of directionality gained within the DI framework. In other words, DI zones hyper-T1 and hyper-T2 (see Eq. 4 and 5 above, and Figure 4) both indicate large relative increases in allopolyploid niche breadth and separation in niche centroid position away from progenitors but are distinguished by the directionality of those changes towards the lower (–) or the higher (+) values observed in the progenitors. The same added resolution and directionality distinguish hyper-T2 from hyper-T3, hyper-T3 from hyper-T4, and hyper-T4 from hyper-T1 across one or both DI axes (DI_C and/or DI_B) (Figure 4). Additionally, there are four hypo-transgression zones (hypo-T1–4) that mirror the four hyper-transgression zones. These zones of

hypo-transgression are simply defined as an allopolyploid with all $DI_{C,B}$ values beyond progenitors (beyond $\pm 1 DI_{C,B}$) but not exceeding the progenitor difference ($\pm 3 DI_{C,B}$).

Intermediacy: Intermediacy is intuitively conceptualized as a characteristic sharing similarity with both progenitors and is perhaps the most common null hypothesis in polyploid and hybrid biology (e.g., Winge, 1932). Here, I define the concept of intermediacy (H_0) to mean that a polyploid exhibits values strictly between the progenitor values where “perfect” intermediacy is the arithmetic mean of progenitor centroids and breadths ($\bar{D}_{C,B}$). With respect to niche occupancy, this means that both the breadth and centroid position should be between parental values such that:

$$D_{1,B} \leq PP_B \leq D_{2,B} \quad (8)$$

and

$$D_{1,C} \leq PP_C \leq D_{2,C} \quad (9)$$

where PP_B is the niche breadth of the allopolyploid and PP_C is the allopolyploid centroid. $D_{1,C}$ is the centroid of progenitor 1, $D_{1,B}$ is the breadth of progenitor 1, and $D_{2,C}$ and $D_{2,B}$ are the centroid and breadth of progenitor 2, respectively. In this example, parental data are arbitrarily assigned from smallest (1) to largest (2).

In the DI framework, intermediacy (H_0) is defined as an allopolyploid niche breadth and centroid between progenitor values such that:

$$-1 \leq DI_C \leq 1 \quad (10)$$

and

$$-1 \leq DI_B \leq 1 \quad (11)$$

where DI_C and DI_B are the niche centroid and niche breadth, respectively, -1 and 1 on both axes represent progenitor values, and 0 indicates “perfect” intermediate values ($\bar{D}_{C,B}$) (Figure 4).

Contraction (C): Contraction (C) has been defined as a polyploid niche breadth that is smaller than a threshold that varies by study, but all studies share the conceptual agreement that the polyploid niche breadth shows a size reduction. Here, I generally constrain contraction to require:

$$D_{1,B} > PP_B \quad (12)$$

and

$$D_{2,B} > PP_B \quad (13)$$

where PP_B is the niche breadth of the allopolyploid, $D_{1,B}$ is the breadth of progenitor 1, and $D_{2,B}$ is the breadth of progenitor 2.

In the DI framework, any movement on the DI_B axis represents a change in the allopolyploid's niche breadth relative to the parental niche breadths, where -1 is equal to the breadth of the parent with the smaller breadth, and 1 is equal to the breadth of the parent with the larger breadth (see Figures 3, 4). Thus, assessing niche contraction (C) is very simple in the DI framework and occurs in negative DI_B values. Changes along the DI_C axis would indicate other changes in niche occupancy not invoked by the concept of niche contraction per se (see other ecological outcomes below). Therefore, I generally define niche contraction as an allopolyploid with an intermediate DI_C and negative DI_B and further distinguish between moderate degrees of contraction (hypo-C) and extreme degrees of contraction (hyper-C) by imposing a threshold of $-3 DI_B$.

hyper-contraction (hyper-C): An allopolyploid with an intermediate niche centroid (DI_C) and DI niche breadth (DI_B) less than -3 such that:

$$-1 \leq DI_C \leq 1 \quad (14)$$

and

$$DI_B < -3 \quad (15)$$

where DI_C represents the allopolyploid niche centroid in the DI framework and DI_B represents the allopolyploid niche breadth in the DI framework (Figure 4). This definition of hyper-contraction (hyper-C) represents a reduction in an allopolyploid niche breadth by a magnitude greater than the difference between progenitor niche breadths.

hypo-contraction (hypo-C): An allopolyploid with an intermediate niche centroid (DI_C) and DI niche breadth (DI_B) less than -1 and greater than or equal to -3 such that:

$$-1 \leq DI_C \leq 1 \quad (16)$$

and

$$-1 > DI_B \geq -3 \quad (17)$$

This definition of hypo-contraction (hypo-C) represents a reduction in an allopolyploid niche breadth but not by more than the difference between progenitor niche breadths.

Expansion (E): Expansion (E) is characterized as an allopolyploid niche breadth that is larger than a threshold that varies by study, but all share the conceptual agreement that the polyploid niche breadth shows an increase in size. Here, I generally constrain expansion to require:

$$D_{1,B} < PP_B \quad (18)$$

and

$$D_{2,B} < PP_B \quad (19)$$

where PP_B is the niche breadth of the allopolyploid, $D_{1,B}$ is the breadth of progenitor 1, and $D_{2,B}$ is the breadth of progenitor 2.

In the DI framework, any movement on the DI_B axis represents a change in the allopolyploid's niche breadth relative to the parental niche breadths. Thus, niche expansion (E) in the DI framework occurs in positive DI_B values (Figure 4). Changes along the DI_C axis would indicate other changes in niche occupancy (i.e., centroid position) not invoked by the concept of niche expansion per se (see other ecological outcomes below). Therefore, I generally define niche expansion as an allopolyploid with an intermediate DI_C and positive DI_B , and further distinguish between moderate degrees of expansion (hypo-E) and extreme degrees of expansion (hyper-E) by imposing a threshold of 3 DI_B .

hyper-expansion (hyper-E): An allopolyploid with an intermediate niche centroid (DI_C) and niche breadth (DI_B) greater than 3 such that:

$$-1 \leq DI_C \leq 1 \quad (20)$$

and

$$DI_B > 3 \quad (21)$$

where D_C represents the allopolyploid niche centroid in the DI framework and DI_B represents the allopolyploid niche breadth in the DI framework (Figure 4). This definition of hyper-expansion represents an increase in polyploid niche breadth by a magnitude greater than the difference between parental niche breadths.

hypo-expansion (hypo-E): An allopolyploid with an intermediate niche centroid (DI_C) and niche breadth (DI_B) greater than 1 and less than or equal to 3 such that

$$-1 \leq DI_C \leq 1 \quad (22)$$

and

$$1 < DI_B \leq 3 \quad (23)$$

This definition of hypo-expansion (hypo-E) represents an increase in polyploid niche breadth but not by more than the difference between parental niche breadths.

Niche shift: The degree of separation in niche occupancy relative to progenitor niches is thought to impact the ecological success of newly formed polyploids (e.g., Levin, 1975; Fowler and Levin, 1984, 2016). Many studies have characterized “niche shifts,” and there are nearly as many

interpretations and definitions of what constitutes a niche shift as there are studies. However, the concept of a niche shift is more strongly tied to changes in centroid position than to changes in breadth (Guisan et al., 2014). This is because changes in centroid can occur without any change in breadth, while changes in breadth will almost always change the centroid—excepting, for example, the case of a change in breadth that is symmetrical and equal in magnitude across all dimensions. Furthermore, the DI framework already incorporates and classifies changes in breadth independently of significant shifts in centroid position (i.e., expansion and contraction). Therefore, I generally define a niche shift (NS^\pm) as a relative change in centroid position beyond parental niche centroids, and without large changes in niche breadth, such that:

$$D_{1,B} \leq PP_B \leq D_{2,B} \quad (24)$$

and

$$D_{1,C} > PP_C \mid D_{2,C} < PP_C \quad (25)$$

where PP_B is the breadth of the polyploid and PP_C is the allopolyploid centroid. $D_{1,C}$ is the centroid of progenitor 1, $D_{1,B}$ is the breadth of progenitor 1, and $D_{2,C}$ and $D_{2,B}$ are the centroid and breadth, respectively, of progenitor 2. Progenitor data are arbitrarily assigned from smallest (1) to largest (2).

In the DI framework, any movement on the DI_C axis represents a shift in the allopolyploid's centroid relative to the progenitor centroids. Thus, niche shifts (NS) in the DI framework are not a single category, but rather comprise two directional zones that occur in large positive or negative DI_C values. Changes along the DI_B axis would indicate other changes in niche occupancy (i.e., breadth) not strictly invoked by the concept of a niche shift as defined here (see other biological outcomes above). Therefore, I generally define NS as an allopolyploid with an intermediate DI_B and positive or negative DI_C and further distinguish between moderate degrees of niche shift (hypo-NS) and extreme degrees of niche shift (hyper-NS) by imposing a threshold of ± 3 DI_C .

hyper-niche shift (hyper-NS[±]): A shift in allopolyploid niche position (as indicated by the centroid) by a magnitude greater than the difference between progenitor niche centroids in either the positive (NS^+) or negative (NS^-) direction, such that:

$$-1 \leq DI_B \leq 1 \quad (26)$$

and

$$DI_C > 3 \mid DI_C < -3 \quad (27)$$

where DI_C represents the allopolyploid niche centroid position and DI_B represents the allopolyploid niche breadth in the DI framework (Figure 4).

hypo-niche shift (hypo-NS⁺): A polyploid with an intermediate niche breadth (DI_B) and niche centroid (DI_C) greater than 1 and less than or equal to 3 such that:

$$-1 \leq DI_B \leq 1 \quad (28)$$

and

$$1 < DI_C \leq 3 \quad (29)$$

hypo-niche shift (hypo-NS⁻): A polyploid with an intermediate niche breadth (DI_B) and niche centroid (DI_C) less than -1 and greater than or equal to -3 such that

$$-1 \leq DI_B \leq 1 \quad (30)$$

and

$$-1 > DI_C \geq -3 \quad (31)$$

A key benefit of using the DI framework is that shifts in niche (or any continuous trait) are directional with explicit reference to progenitors, rather than qualitative and/or directionally ambiguous. See below for examples and a discussion of current limitations.

Although I focused on the application of these concepts in an ecological context, recall that the above classifications can be made with any continuous data as inputs and need not be ecological data.

METHODS

Examples of using the DI framework for trait divergence

To show how organismal trait data can be used inside the DI framework, I gathered trait data from multiple studies. First, I gathered data from Mitchell et al. (2019), which examined the evolution of morphological and functional traits in a sunflower hybrid over eight generations in a common garden. Briefly, Mitchell et al. (2019) planted common gardens of *Helianthus annuus* L. subsp. *annuus* and a hybrid backcross that was experimentally formed by crossing *H. annuus* subsp. *annuus* with the resynthesized hybrid between *H. annuus* subsp. *annuus* and *H. debilis* Nutt. Over eight generations, Mitchell et al. (2019) measured several traits related to life history, reproduction, growth, and function in the backcross hybrid and *H. annuus* subsp. *annuus*, and performed Bayesian linear regression models to estimate the evolutionary change in phenotypic values of traits through time. I used data reported in Mitchell et al. (2019) from the backcross and *H. annuus* subsp. *annuus*, and trait data from the original common garden studies that included the other progenitor, *H. debilis* (Whitney et al., 2006, 2010) (see Table 1 for a list of traits). I used reported trait means (raw data were not available) to

TABLE 1 A list of traits gathered from studies to demonstrate the divergence index (DI) framework (see Figures 5 and 6).

Trait code	Description	Units	Studies ^a	Figure
CNratio	Leaf carbon to nitrogen ratio	unitless ratio	1,2,3	5
DaysToBud	Bud initiation time	days	1,2,3	5
GlandDens	Glandular trichome density	mm ²	1,2,3	5
HairDens	Non-glandular trichome density	mm ²	1,2,3	5
HtLow	Height of lowest branch	cm	1,2,3	5
Longevity	Lifespan of plant	days	1,2,3	5
RelBrDiam	Relative branch diameter	unitless	1,2,3	5
SLA	Specific leaf area	cm ² /g	1,2,3	5
SMT	Seed maturation time	days	1,2,3	5
iWUE	Integrated water use efficiency ($\delta^{13}C$)	‰	1,2,3	5
Plant Height	Height of terminal apex	cm	4	6
PhiPSII	Photochemical efficiency of PSII	unitless ratio	4	6
Seed Number	Number of seeds produced by plant	No.	4	6

^a1, Mitchell et al. (2019); 2, Whitney et al. (2006); 3, Whitney et al. (2010); 4, Jordon-Thaden et al. (2023).

calculate a DI_C for each generation. In other words, I used *H. annuus* subsp. *annuus* and the F_1 hybrid between *H. annuus* subsp. *annuus* and *H. debilis* as “progenitors” (DD_1 and DD_2), and then calculated a DI_C for each generation of the backcrossed hybrid (PP_{gen1} , PP_{gen2} , etc.). I then calculated the difference in DI_C between generations eight and one (ΔDI_C). I performed a standard major axis regression using the *sma()* function from the *smatr* package in R (Warton et al., 2012) between the evolution slope reported for each trait in Mitchell et al. (2019) and the ΔDI_C of each trait (Figure 5). I used the *ggplot2* package to plot the data (Wickham et al., 2016). The sunflower data can be found in Appendix S1, and DI values can be found in Appendix S2. The DI values, evolution slope data, and data plotted in Figure 5 can be found in Appendix S3. The code to reproduce results and figures can be found in Appendix S4.

To further demonstrate using the DI framework for the unification of data from different species and variables that differ in scale, I gathered data from Jordon-Thaden et al. (2023) on the phenotypic divergence between natural allopolyploids and synthetic allopolyploids of North American *Tragopogon* L. Briefly, Jordon-Thaden et al. (2023) planted a common garden of natural diploid progenitors (*Tragopogon dubius* Scop.,

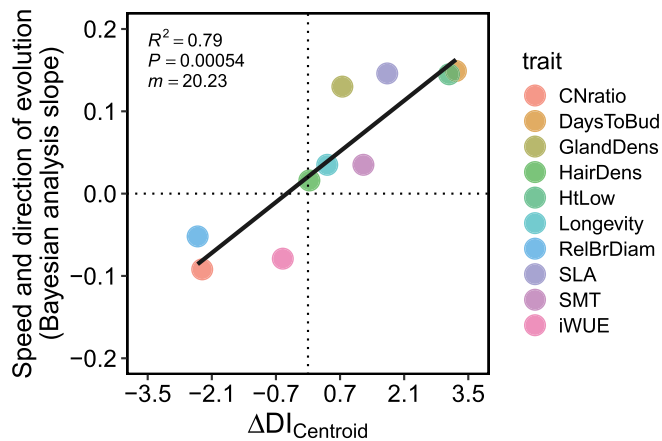


FIGURE 5 Simple divergence index (DI) values recapitulate trends from complex model output. Data are from Mitchell et al. (2019) and represent 10 traits measured in a common garden of over eight generations. The x -axis is ΔDI_C and indicates the change in DI_C values over eight generations. The means and standard deviations of raw data were used as inputs to calculate generational DI_C values. The y -axis is the slope of evolution for each trait from the Bayesian models performed in Mitchell et al. (2019). See Table 1 for a list of trait descriptions.

T. porrifolius L., and *T. pratensis* L.) and allopolyploids (*T. mirus* Ownbey and *T. miscellus* G. B. Ownbey), and re-synthesized allopolyploids (using colchicine). The authors measured a suite of morphological and growth trait and one physiological trait (see Table 1). For each trait and species, I extracted the standard deviation as a metric of breadth and used the *density* function to smooth the distribution of raw trait data before extracting the 50th percentile as a metric of the centroid. I could use the density smoothing approach to estimate a centroid because the raw data were made available, rather than being reported only as trait means. To quantify and visualize the difference in traits with very different scales between natural and synthetic allopolyploids, I calculated DI_C and DI_B values for plant height, photochemical efficiency (PhiPSII), and seed number for the natural and synthetic allopolyploids *T. mirus* and *T. miscellus* (Figure 6). The scale of the raw data for PhiPSII is on the order of 0.1, while the scale of the raw data for seed number is on the order of 100. I used the ggplot2 package to plot the data (Wickham et al., 2016). The raw *Tragopogon* data can be found in Appendix S5, and DI values can be found in Appendix S6. The code to reproduce Figure 6 can be found in Appendix S7.

Example of using the DI framework for niche divergence

To show how ecological data can be used inside the DI framework, I gathered data on the distribution of four extant allopolyploid complexes. The ecological patterns of these four complexes have been previously characterized by Marchant et al. (2016) as exhibiting niche contraction (*Cystopteris bulbifera* (L.) Bernh., *C. protrusa* (Weath.) Blasdell, *C. tennesseensis* Shaver; Trio 1), niche intermediacy

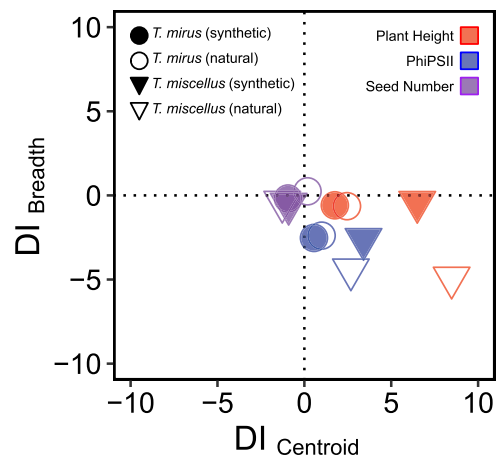


FIGURE 6 The divergence index (DI) framework standardizes traits that differ in scale by multiple orders of magnitude. Data are from Jordan-Thaden et al. (2023) and represent natural (open shapes) and synthetic (filled shapes) *Tragopogon mirus* (circles) and *T. miscellus* (triangle). Red indicates plant height, blue indicates photochemical efficiency (PhiPSII), and purple indicates seed number. Raw data were used as inputs to calculate DI_C . See Table 1.

(*Dryopteris celsa* (W. Palmer) Knowlt., T. S. Palmer & Pollard ex Small, *D. goldieana* (Hook. ex Goldie) A. Gray, *D. ludoviciana* (Kunze) Small; Trio 2), niche expansion (*Polypodium amorphum* Suksd., *P. glycyrrhiza* D. C. Eaton, *P. hesperium* Maxon; Trio 3), and niche transgression (*Polypodium amorphum*, *P. saximontanum* Windham, *P. sibiricum* Sipliv.; Trio 4).

Geographic occurrence records

I downloaded geographic occurrence records for these species from GBIF (<https://www.gbif.org/>), iDigBio (<https://www.idigbio.org/>), and the Atlas of Living Australia (<https://www.ala.org.au/>) using the spocc R package (Chamberlain et al., 2016). After removing duplicate records and rows with missing data, the remaining records were cleaned using the CoordinateCleaner R package (Zizka et al., 2019) to identify problematic records, including those that fell into the ocean(s) using Natural Earth medium-resolution (1:50 m) land mass boundaries; records with null, identical, or “0” values for latitude and longitude; and records with coordinates associated with capital cities, political centroids, or known institutions. The *coordError* function from the rangeBuilder R package (<https://github.com/ptitle/rangeBuilder>) was used to calculate the maximum potential error associated with the reported coordinate precision, and I removed occurrence records with a potential error larger than 10 km. Geographic outliers were identified and removed using the *cc_outl* function in the CoordinateCleaner R package. The *FilterByProximity* function in rangeBuilder was used to filter records to a spatial density of ≥ 1 km between points. Finally, isolation forests were applied using the solitude R package (Liu et al., 2012)

to identify environmental outliers. Within each species, occurrence records with an anomaly score greater than 0.7 in climate space were removed (with 0 indicating a point that is difficult to isolate from the group and 1 indicating a point that is very easy to isolate from the group) (Liu et al., 2008, 2012).

Environmental data

Environmental data were gathered from many different sources including CHELSA, SoilGrids, SoilTemp, and MODIS (Sexton et al., 2013; Karger et al., 2017; Lembrechts et al., 2020; Poggio et al., 2021). In total, 35 environmental variables were selected to represent axes of temperature (e.g., air and soil temperature), water availability (e.g., precipitation and soil moisture), atmospheric demand for water (e.g., potential evapotranspiration and vapor pressure deficit), light availability (e.g., leaf area index, solar radiation, cloud cover), and soil fertility (e.g., nitrogen, phosphorus, organic carbon) (a complete list of variables and their sources can be found in Appendix S8). Environmental data were extracted from global rasters using the point-sampling-tool in QGIS (<https://www.qgis.org/>). All 35 environmental variables were rescaled between 1 and 100 using the global minimum and maximum raster values (the data were not centered). Next, variables were reverse scaled as needed so that higher values generally represented more favorable conditions and lower numbers were less favorable. For example, vapor pressure deficit was reverse scaled so that low values indicate dry air and high values indicate moist air, and leaf area index (a proxy for shade) was reverse scaled so that low values indicate less light and higher values indicate more light. Conversely, variables such as precipitation, soil moisture, air and soil temperature, soil phosphorus, soil nitrogen, and soil organic carbon were not reverse scaled. This directional alignment of the input environmental variables is important to maintain the function of the positive and negative range of the DI_C axis when using the DI framework with many variables. This alignment and orientation are not necessary when using the DI framework with one variable. See “Current limitations” for advice when alignment is not achievable.

Quantifying niche centroid and breadth

For each allopolyploid complex, I performed a principal component analysis (PCA) on their environmental data using the *prcomp* function in R. For each species and principal component axis, I extracted the standard deviation as a metric of variation or breadth (DI_B). I used the *density* function to smooth the distribution of data (PCA scores) before extracting the 50th percentile as a metric of central tendency or centroid (DI_C). Only the principal component axes that collectively represent significant variation in the dataset were retained for further analyses (usually five to

seven axes that explain 80–90% of total variation; see Appendix S8). I calculated a weighted mean centroid for each species (DI_C) using the centroid values for each principal component axis weighted by the variance explained by the axis. Similarly, I calculated a weighted mean breadth for each species (DI_B) using the breadth values for each principal component axis weighted by the standard deviation of the principal component axis. DI values for each complex were visualized using the *ggplot2* R package (Wickham et al., 2016). See Appendix S8 for PCA data and Appendix S9 for associated code.

All analyses were performed in the R coding environment (v4.3) unless otherwise noted. For all R packages and their functions, default parameters were used unless otherwise noted.

RESULTS

Example of using the DI framework for trait divergence

The standard major axis regression revealed a strong relationship between the evolution slope and ΔDI_C from multi-generational trait data of a hybrid sunflower (Whitney et al., 2006, 2010; Mitchell et al., 2019) (Figure 5; $P = 0.00054$, $R^2 = 0.79$, $m = 20.23$). The phenotypic outcomes of seed number, PhiPSII, and plant height in natural and synthetic *Tragopogon* can be seen in Figure 6.

Example of using the DI framework for niche divergence

Trio 1 (*Cystopteris tennesseensis* complex) fell into the hyper-contraction zone within the DI framework, with a DI_C of 0.16 and DI_B of -3.29 (see Eq. 14 and 15; Figure 7).

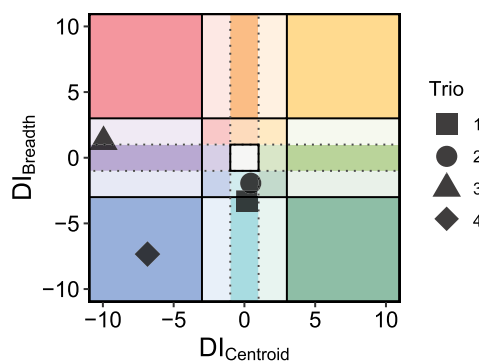


FIGURE 7 The ecological outcomes of four allopolyploid complexes: Trio 1, *Cystopteris tennesseensis* complex (square); Trio 2, *Dryopteris celsa* complex (circle); Trio 3, *Polypodium hesperium* complex (triangle); and Trio 4, *Polypodium saximontanum* complex (diamond). The inner-most square box with solid lines represents the intermediacy zone. Dotted lines are drawn at -1 and $+1$ on both axes and solid lines are drawn at -3 and $+3$ on both axes.

Trio 2 (*Dryopteris celsa* complex) fell into the hypo-contraction zone within the DI framework, with a DI_C of 0.45 and DI_B of -1.94 (see Eq. 16 and 17). Trio 3 (*Polypodium hesperium* complex) exhibited a slight increase in niche breadth relative to parental taxa ($DI_B = 1.12$) and a large shift in niche centroid beyond parental centroids ($DI_C = -9.95$). Trio 4 (*Polypodium saximontanum* complex) fell into the hyper-transgression zone and specifically the hyper-T3 zone within the DI framework, with a DI_C of -6.89 and DI_B of -7.52 (see Eq. 6; Figure 7).

DISCUSSION

To show how the DI framework can be used with empirical data to capture trends and gain insight to hybrid and polyploid biology, I used empirical data from multiple sources that utilized one or both of the $DI_{C,B}$ axes. For example, Mitchell et al. (2019) found that hybridization accelerated rates of phenotypic evolution in a backcross hybrid of sunflower. Using trait data in the DI framework, I found that the slopes of evolutionary change from the Bayesian models performed by Mitchell et al. (2019) were highly correlated with the change in DI_C values of 10 traits over eight generations in the common garden (ΔDI_C) ($R^2 = 0.79$; Figure 5). Mitchell et al. (2019) found a similar pattern between the slope of evolutionary change and the difference in predicted mean of the normalized trait data from the Bayesian models (see figure 4a in Mitchell et al., 2019). Here, I used the means of the raw data in the DI framework to recapitulate this pattern of trait evolution using the DI framework (Figure 5). This analysis demonstrates that the DI framework can be used with many different traits alongside and in place of the output from complex models, despite the DI framework being remarkably simple in comparison and without normalization of the underlying data (i.e., without setting mean to zero and standard deviation to one). Similarly, I used data from Jordon-Thaden et al. (2023) to show that multiple traits from multiple species can be easily quantified and visualized in the same plot by using the DI framework. Jordon-Thaden et al. (2023) investigated potential differences in biological outcomes between natural and synthetic allopolyploids of *Tragopogon*. The three variables depicted in Figure 6 (i.e., PhiPSII, plant height, and seed number) differ in scale by three orders of magnitude (0.1 to 100; e.g., PhiPSII to seed number). Thus, the DI framework offers a flexible and intuitive tool to unify data from different traits that also differ in scale for investigations and characterization of hybrid and polyploid biodiversity.

Translating empirical data for allopolyploid niche occupancy that has already been characterized using other methods revealed that the DI framework captures many of the same trends but with some important differences. Two of the four Trios (Trio 1 and 4) occurred in DI categories that are nominally analogous to other methods of categorization. Specifically, Trio 1 has been previously

characterized as exhibiting niche contraction, and this is exactly how it is characterized within the DI framework (Figure 7) but with additional information on the magnitude of the niche contraction conferred by the hyper-contraction zone criteria. Similarly, Trio 4 has been previously characterized as exhibiting niche transgression, and this complex also occurs within a DI transgression zone. However, by using the DI framework, we can see that this complex (Trio 4) achieved niche transgression with significant and concurrent changes in niche centroid and reduction in niche breadth (hyper-transgression zone 3; see Eq. 6). Characterizing the ecological outcomes of Trios 2 and 3 similarly added new insight to the mechanisms that underly patterns of niche differentiation in these allopolyploid complexes. For example, Trio 3 has been previously characterized as exhibiting niche expansion and this Trio did in fact exhibit a larger niche breadth than both progenitors ($DI_B = 1.12$). However, the DI framework also revealed that Trio 3 surpassed the benchmark of $\pm 3 DI_C$ ($DI_C = -9.95$), indicating that this allopolyploid achieved a niche shift beyond parental taxa that is also greater than the difference between progenitor niche centroids, thus achieving a hyper-NS⁻ (see Eq. 26 and 27) (Figure 7). The added insight into changes in niche centroid and breadth that underly niche occupancy patterns opens new doors for understanding the physiological mechanisms that explain species ecology and distribution.

The allopolyploid complex in Trio 1 has been previously characterized as exhibiting niche intermediacy and, indeed, the DI_C is between -1 and 1 ; however, the niche breadth exhibited by this allopolyploid is moderately smaller than the smallest parental breadth ($DI_B = -1.94$), indicating hypo-contraction (see Eq. 16 and 17; Figure 7) rather than intermediacy. This difference in categorical characterization is likely because other methods have incorporated thresholds of niche overlap into their definitions of these categories. In contrast, the DI framework dually quantifies changes in centroid and breadth independently of overlap. Niche overlap is an important metric of ecology that offers information that is also theoretically independent of the information gained with the DI framework. Consider the hypothetical case where a polyploid is perfectly intermediate between its progenitors in every way but does not overlap with either progenitor because the progenitors are sufficiently distinct (i.e., progenitor niches are sufficiently separated that the polyploid occupies an intermediate niche without overlap). Similarly, metrics of niche overlap would cease to be useful in the case of a highly divergent polyploid that also did not overlap with either progenitor. Thus, the DI framework is highly complementary to metrics of niche overlap (Schoener, 1970; Warren et al., 2008; Broennimann et al., 2012; Blonder et al., 2018; Brown et al., 2020; Lu et al., 2021), and using metrics of niche overlap in tandem with the DI framework has the potential to further characterize patterns in polyploid and hybrid biology. See below, under “Current limitations,” for further discussion of overlap.

The DI framework can be used in univariate and multivariate analyses and will always retain the same interpretation of hybrid and/or allopolyploid characteristics to parental differences (i.e., one unit in DI space is always equal to one-half progenitor difference, and benchmark values such as $DI \pm 1$ and $DI \pm 3$ always have the same definition and interpretation regardless of inputs or dimensionality). Therefore, the DI framework produces standardized output that is directly comparable within and across studies, as well as across different variable types. However, to maximize comparability across studies, multivariate centroid and breadth data can each be summarized in one dimension. For example, when calculating the volume (i.e., breadth) of an n -dimensional hypervolume, the units are reported in standard deviations to the n^{th} power (n being the number of dimensions) (Blonder et al., 2014, 2018) and the n^{th} root should be taken as the volume (DI_B). This is because the DI equations are simply distances and ratios, and adding exponents to data entities alters the relationship between entities (e.g., PP, D_1 , D_2 , numerators, and denominators) without altering the defined benchmarks in DI space (i.e., $DI \pm 1$ and $DI \pm 3$ always have the same definition). Alternatively, a weighted mean can be calculated to characterize a multivariate centroid and a multivariate breadth (as I have done here). The convention of calculating a weighted mean from many dimensions is common in hybrid biology (Hatheway, 1962) and is also used by other popular methods to characterize niche position and breadth (e.g., ecospat R package; Di Cola et al., 2017). Another benefit of characterizing niche characteristics in one dimension (e.g., a weighted mean) is the added ease with which DI values can then be used with established statistical tests (e.g., GLMM, perMANOVA), as well as in simulations for hypothesis testing.

The simplicity of the DI framework also makes it among the most accessible methods in polyploid biodiversity research. I advocate that studies of allopolyploid and hybrid biology report the $DI_{C,B}$ values for any appropriate data (e.g., environmental, ecological, morphological, physiological, biochemical, genetic) in their results so that they can be compared across studies and interpreted in a unified and coherent framework. Overall, the introduction of the DI framework generates new research opportunities in comparative biology and can be used to reconcile key issues that have limited the synthesis of ecological theory and empirical data in polyploid and hybrid biodiversity research.

Current limitations

It is important to note that the DI framework requires that the input variables in the denominator (e.g., D_1 and D_2 ; see Eq. 1) are not equal, as this would produce infinite DI values. In practice, this should be rare, and when/if it occurs, it likely reflects limitations in the quality of the data and/or the underlying methodology. In the hypothetical case of measuring plant height across numerous individuals,

the resolution of measurement plays a crucial role. When measurements are taken on a small scale, such as in millimeters, the likelihood of plants having the same height is very low. Conversely, if measurements are made in meters, the potential for capturing meaningful variation in plant height diminishes, leading to identical measurements. Therefore, obtaining high-quality data with appropriate precision (i.e., significant figures) is essential for optimal results within the DI framework. Similarly, very small differences between progenitor values increase the potential for extreme (positive or negative) DI values. Thus, the DI should only be applied when progenitor values are meaningfully differentiated, which can be determined by expert knowledge of species biology and/or an appropriate statistical test performed on progenitor data. In general, the potential for small denominators to produce large DI values may decrease with the number of input variables because of the decreasing probability of two species being very similar across an increasing number of axes.

In the example presented in this manuscript of using the DI framework to characterize niche divergence, I aligned the input variables to make interpretation of the DI_C axis easier and more meaningful. If multivariate data cannot be reasonably aligned in this manner, then individual variables can be used and the absolute value can be taken in the numerator of the equation for DI_C (i.e., using Eq. 2 for centroids). This would remove the negative values in DI_C and produce a positive DI_C scale that quantifies the distance between polyploid and progenitor centroids, where $DI_C = 0$ is intermediate and 1 is either parent. Note that this approach may be useful as a more general metric of centroid divergence but sacrifices directionality with respect to progenitors' centroid values and is therefore a more opaque metric of centroid divergence.

Another limitation to using the two-dimensional DI framework for polyploid biology is that it is easier to parameterize with interspecific hybrids and allopolyploids than with autopolyploids and intra-specific hybrids because of the difficulty in defining two distinct "progenitors" (D_1 and D_2) required to calculate $DI_{C,B}$ values. However, several studies have examined intra-specific hybrids and/or autopolyploids between distinct populations or ecotypes of the same species (e.g., Lowry et al., 2009), and these data would be usable in the DI framework.

Although example data were used to characterize niche divergence within the DI framework using centroids and breadths, quantifying overlap may be important for some biological questions. Niche occupancy may be such a case, and some combinations of species centroid and breadths may violate the expected pattern of overlap in some ecological outcomes. For example, there may be some expectation of complete overlap in the DI zone of intermediacy, but a hybrid that has a centroid and a breadth that are both between the observed values in progenitors may still only exhibit partial overlap when the hybrid centroid is close to the lower progenitor extreme, its breadth is close to the higher progenitor extreme (e.g., DI_C close to -1 , and

DI_B close to 1), and the difference between progenitor breadths is relatively large. Similarly, there may be an expectation of no overlap in DI zones of transgression and niche shift, yet a hybrid with a very large distance from parental centroids (beyond ± 3) could still be within parental ranges if a progenitor breadth is sufficiently large, the hybrid's breadth is sufficiently small, and the distance between progenitor centroids is small. I performed simulations with univariate input data to test how often these two cases of unexpected patterns of overlap might occur. I found that unexpected patterns of overlap are only possible in 13 of the 25 DI zones and all cases of unexpected patterns of overlap (except intermediacy) occurred with negative DI_B values, which indicates that some degree of breadth contraction (as defined in DI space) is a prerequisite for unexpected overlap. Within the 13 zones that had any observation of unexpected overlap, the lowest (non-zero) occurrence was 0.58% and the highest was 3.2%. See Appendix S10 for full details on these simulations. The results of the simulations suggest that the potential for an unexpected pattern of overlap for a given DI zone is relatively rare with univariate data. Moreover, this outcome may be even less common as the number of input variables increases because of the overwhelming probability of additional axes to exhibit the expected pattern of overlap and thus for an observation (hybrid or polyploid complex) to be classified correctly with the expected pattern of overlap (see Holt, 2009). When using the DI to characterize niche divergence, a metric of niche overlap can be used to complement DI values and provide additional insight. Measuring the overlap of data distributions is not currently possible in the DI framework presented here, and future work will focus on developing a robust metric of overlap that can be natively calculated within the DI framework (Krieg, unpublished data).

The DI framework is designed to work with any continuous variable and cannot directly be used with discrete or categorical input data. One option to incorporate data of mixed types (e.g., continuous and discrete) would be first to transform data, e.g., using a Gower transformation (Gower, 1971). However, the properties of Gower-transformed data within the DI framework have not been tested and such an approach warrants caution and careful exploration. The inclusion and exploration of genetic data in the DI framework is beyond the scope of this paper; however, genetic data could be used in the DI framework after transformations designed for genetic data such as the polyploidy index (Wang et al., 2019) or hybrid index (Anderson, 1953; Goulet et al., 2017; Bailey, 2024).

AUTHOR CONTRIBUTIONS

C.P.K. is the sole author of this manuscript.

ACKNOWLEDGMENTS

I am very grateful to Dr. Duncan Smith, Dr. Niki Patel, Dr. Nisa Karimi, and Dr. Michelle Gaynor for their helpful discussions and comments on earlier versions of this

manuscript. I'm also grateful to Dr. Duncan Smith for their contributions to Video S1. I am additionally thankful for the constructive feedback from two reviewers, the Reviewing Editor, the Associate Editor, and the Managing Editor. This research was supported by the U.S. National Science Foundation (DBI-1907033 and IOS-2243970 to C.P.K.).

CONFLICT OF INTEREST STATEMENT

Christopher P. Krieg is an associate editor of *Applications in Plant Sciences*, but took no part in the peer review and decision-making processes for this paper.

DATA AVAILABILITY STATEMENT

All data are available in the Supporting Information.

ORCID

Christopher P. Krieg  <http://orcid.org/0000-0002-9797-9280>

REFERENCES

- Anderson, E. 1953. Introgressive hybridization. *Biological Reviews of the Cambridge Philosophical Society* 28: 280–307.
- Anderson, E. C., and E. A. Thompson. 2002. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* 160: 1217–1229.
- Bailey, R. I. 2024. Bayesian hybrid index and genomic cline estimation with the R package gghybrid. *Molecular Ecology Resources* 24: e13910.
- Baniaga, A. E., H. E. Marx, N. Arrigo, and M. S. Barker. 2020. Polyploid plants have faster rates of multivariate niche differentiation than their diploid relatives. *Ecology Letters* 23: 68–78.
- Blonder, B., C. Lamanna, C. Violle, and B. J. Enquist. 2014. The n -dimensional hypervolume. *Global Ecology and Biogeography* 23: 595–609.
- Blonder, B., C. B. Morrow, B. Maitner, D. J. Harris, C. Lamanna, C. Violle, B. J. Enquist, and A. J. Kerkhoff. 2018. New approaches for delineating n -dimensional hypervolumes. *Methods in Ecology and Evolution* 9: 305–319.
- Bombliès, K. 2020. When everything changes at once: Finding a new normal after genome duplication. *Proceedings of the Royal Society B. Biological Sciences* 287: e20202154.
- Brochmann, C., A. K. Brysting, I. G. Alsos, L. Borgen, H. H. Grundt, A.-C. Scheen, and R. Elven. 2004. Polyploidy in arctic plants. *Biological Journal of the Linnean Society* 82: 521–536.
- Broennimann, O., M. C. Fitzpatrick, P. B. Pearman, B. Petitpierre, L. Pellissier, N. G. Yoccoz, W. Thuiller, et al. 2012. Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography* 21: 481–497.
- Brown, M. J. M., B. R. Holland, and G. J. Jordan. 2020. hyperoverlap: Detecting biological overlap in n -dimensional space. *Methods in Ecology and Evolution* 11: 513–523.
- Buggs, R. J. A., P. S. Soltis, and D. E. Soltis. 2009. Does hybridization between divergent progenitors drive whole-genome duplication? *Molecular Ecology* 18: 3334–3339.
- Chamberlain, S., K. Ram, and T. Hart. 2016. spocc: R interface to many species occurrence data sources. *R package version* 50. Website <https://github.com/ropensci/spocc> [accessed 2 August 2024].
- Clausen, J., D. Keck, and W. M. Hiesey. 1945. Experimental studies on the nature of species. II. Plant evolution through amphidiploidy and autopolyploidy, with examples from the Madiinae. Carnegie Institution, Washington, D.C., USA.
- Corneillie, S., N. De Storme, R. Van Acker, J. U. Fangel, M. De Bruyne, R. De Rycke, D. Geelen, et al. 2019. Polyploidy affects plant growth and alters cell wall composition. *Plant Physiology* 179: 74–87.

- Di Cola, V., O. Broennimann, B. Petitpierre, F. T. Breiner, M. D'Amen, C. Randin, R. Engler, et al. 2017. ecospat: An R package to support spatial analyses and modeling of species niches and distributions. *Ecography* 40: 774–787.
- Fowler, N. L., and D. A. Levin. 1984. Ecological constraints on the establishment of a novel polyploid in competition with its diploid progenitor. *The American Naturalist* 124: 703–711.
- Fowler, N. L., and D. A. Levin. 2016. Critical factors in the establishment of allopolyploids. *American Journal of Botany* 103: 1236–1251.
- Gaynor, M. L., S. Lim-Hing, and C. M. Mason. 2020. Impact of genome duplication on secondary metabolite composition in non-cultivated species: A systematic meta-analysis. *Annals of Botany* 126: 363–376.
- Glennon, K. L., M. E. Ritchie, and K. A. Segraves. 2014. Evidence for shared broad-scale climatic niches of diploid and polyploid plants. *Ecology Letters* 17: 574–582.
- Goulet, B. E., F. Roda, and R. Hopkins. 2017. Hybridization in plants: Old ideas, new techniques. *Plant Physiology* 173: 65–78.
- Gower, J. C. 1971. A general coefficient of similarity and some of its properties. *Biometrics* 27: 857.
- Grant, V. 1981. Plant speciation. Columbia University Press, New York, New York, USA.
- Guisan, A., B. Petitpierre, O. Broennimann, C. Daehler, and C. Kueffer. 2014. Unifying niche shift studies: Insights from biological invasions. *Trends in Ecology & Evolution* 29: 260–269.
- Hao, G.-Y., M. E. Lucero, S. C. Sanderson, E. H. Zacharias, and N. M. Holbrook. 2013. Polyploidy enhances the occupation of heterogeneous environments through hydraulic related trade-offs in *Atriplex canescens* (Chenopodiaceae). *The New Phytologist* 197: 970–978.
- Harbert, R. S., A. H. D. Brown, and J. J. Doyle. 2014. Climate niche modeling in the perennial *Glycine* (Leguminosae) allopolyploid complex. *American Journal of Botany* 101: 710–721.
- Hatheway, W. H. 1962. A weighted hybrid index. *Evolution* 16: 1–10.
- Hijmans, R. J., T. Gavrilenko, S. Stephenson, J. Bamberg, A. Salas, and D. M. Spooner. 2007. Geographical and environmental range expansion through polyploidy in wild potatoes (*Solanum* section *Petota*). *Global Ecology and Biogeography* 16: 485–495.
- Holt, R. D. 2009. Bringing the Hutchinsonian niche into the 21st century: Ecological and evolutionary perspectives. *Proceedings of the National Academy of Sciences, USA* 106(Suppl 2): 19659–19665.
- Jiao, Y., N. J. Wickett, S. Ayyampalayam, A. S. Chanderali, L. Landherr, P. E. Ralph, L. P. Tomsho, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473: 97–100.
- Jordon-Thaden, I. E., J. P. Spoelhof, L. F. Viccini, J. Combs, F. Gomez, I. Walker, D. E. Soltis, and P. S. Soltis. 2023. Phenotypic trait variation in the North American *Tragopogon* allopolyploid complex. *American Journal of Botany* 110: e16189.
- Karger, D. N., O. Conrad, J. Böhner, T. Kawohl, H. Kreft, R. W. Soria-Auza, N. E. Zimmermann, et al. 2017. Climatologies at high resolution for the earth's land surface areas. *Scientific Data* 4: 170122.
- Lembrechts, J. J., J. Aalto, M. B. Ashcroft, P. De Frenne, M. Kopecký, J. Lenoir, M. Luoto, et al. 2020. SoilTemp: A global database of near-surface temperature. *Global Change Biology* 26: 6616–6629.
- Levin, D. A. 1975. Minority cytotype exclusion in local plant populations. *Taxon* 24: 35–43.
- Levin, D. A. 1983. Polyploidy and novelty in flowering plants. *The American Naturalist* 122: 1–25.
- Liu, F. T., K. M. Ting, and Z.-H. Zhou. 2008. Isolation forest. 2008 Eighth IEEE International Conference on Data Mining, 413–422. IEEE, New York, New York, USA.
- Liu, F. T., K. M. Ting, and Z.-H. Zhou. 2012. Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data* 6: 1–39.
- Losada, J. M., N. Blanco-Moure, A. Fonollá, E. Martínez-Ferri, and J. I. Hormaza. 2023. Hydraulic tradeoffs underlie enhanced performance of polyploid trees under soil water deficit. *Plant Physiology* 192: 1821–1835.
- Lowry, D. B., M. C. Hall, D. E. Salt, and J. H. Willis. 2009. Genetic and physiological basis of adaptive salt tolerance divergence between coastal and inland *Mimulus guttatus*. *The New Phytologist* 183: 776–788.
- Lu, M., K. Winner, and W. Jetz. 2021. A unifying framework for quantifying and comparing n -dimensional hypervolumes. *Methods in Ecology and Evolution* 12: 1953–1968.
- Maherali, H., A. E. Walden, and B. C. Husband. 2009. Genome duplication and the evolution of physiological responses to water stress. *The New Phytologist* 184: 721–731.
- Manzaneda, A. J., P. J. Rey, J. T. Anderson, E. Raskin, C. Weiss-Lehman, and T. Mitchell-Olds. 2015. Natural variation, differentiation, and genetic trade-offs of ecophysiological traits in response to water limitation in *Brachypodium distachyon* and its descendent allotetraploid *B. hybridum* (Poaceae): Ploidy and adaptation in *Brachypodium distachyon*. *Evolution* 69: 2689–2704.
- Marchant, D. B., D. E. Soltis, and P. S. Soltis. 2016. Patterns of abiotic niche shifts in allopolyploids relative to their progenitors. *The New Phytologist* 212: 708–718.
- Mayrose, I., S. H. Zhan, C. J. Rothfels, K. Magnuson-Ford, M. S. Barker, L. H. Rieseberg, and S. P. Otto. 2011. Recently formed polyploid plants diversify at lower rates. *Science* 333: 1257.
- McCulloh, K. A., S. P. Augustine, A. Goke, R. Jordan, C. P. Krieg, K. O'Keefe, and D. D. Smith. 2023. At least it is a dry cold: The global distribution of freeze-thaw and drought stress and the traits that may impart poly-tolerance in conifers. *Tree Physiology* 43: tpac102.
- McIntyre, P. J. 2012. Polyploidy associated with altered and broader ecological niches in the *Claytonia perfoliata* (Portulacaceae) species complex. *American Journal of Botany* 99: 655–662.
- Mitchell, N., G. L. Owens, S. M. Hovick, L. Rieseberg, and K. Whitney. 2019. Hybridization speeds adaptive evolution in an eight-year field experiment. *Scientific Reports* 9: e6746.
- Nieto Feliner, G., J. Casacuberta, and J. F. Wendel. 2020. Genomics of evolutionary novelty in hybrids and polyploids. *Frontiers in Genetics* 11: 792.
- Otto, S. P., and J. Whitton. 2000. Polyploid incidence and evolution. *Annual Review of Genetics* 34: 401–437.
- Parisod, C., and O. Broennimann. 2016. Towards unified hypotheses of the impact of polyploidy on ecological niches. *The New Phytologist* 212: 540–542.
- Paun, O., F. Forest, M. F. Fay, and M. W. Chase. 2009. Hybrid speciation in angiosperms: Parental divergence drives ploidy. *The New Phytologist* 182: 507–518.
- Poggio, L., L. M. de Sousa, N. H. Batjes, G. B. M. Heuvelink, B. Kempen, E. Ribeiro, and D. Rossiter. 2021. SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *Soil* 7: 217–240.
- Ramsey, J. 2011. Polyploidy and ecological adaptation in wild yarrow. *Proceedings of the National Academy of Sciences, USA* 108: 7096–7101.
- Rieseberg, L. H., S. J. E. Baird, and A. M. Desrochers. 1998. Patterns of mating in wild sunflower hybrid zones. *Evolution* 52: 713–726.
- Schoener, T. W. 1970. Nonsynchronous spatial overlap of lizards in patchy habitats. *Ecology* 51: 408–418.
- Sexton, J. O., X.-P. Song, M. Feng, P. Noojipady, A. Anand, C. Huang, D.-H. Kim, et al. 2013. Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. *International Journal of Digital Earth* 6: 427–448.
- Shimizu, K. K. 2022. Robustness and the generalist niche of polyploid species: Genome shock or gradual evolution? *Current Opinion in Plant Biology* 69: 102292.
- Singhal, S., A. B. Roddy, C. DiVittorio, A. Sanchez-Amaya, C. L. Henriquez, C. R. Brodersen, S. Fehlberg, and F. Zapata. 2021. Diversification, disparification and hybridization in the desert shrubs *Encelia*. *The New Phytologist* 230: 1228–1241.
- Soltis, D. E., R. J. A. Buggs, J. J. Doyle, and P. S. Soltis. 2010. What we still don't know about polyploidy. *Taxon* 59: 1387–1403.
- Soltis, D. E., C. J. Visger, and P. S. Soltis. 2014. The polyploidy revolution then... and now: Stebbins revisited. *American Journal of Botany* 101: 1057–1078.

- Soltis, P. S., and D. E. Soltis. 2009. The role of hybridization in plant speciation. *Annual Review of Plant Biology* 60: 561–588.
- Stebbins, G. L. 1971. Chromosomal evolution in higher plants. Edward Arnold, London, United Kingdom.
- Stebbins, G. L. 1985. Polyploidy, hybridization, and the invasion of new habitats. *Annals of the Missouri Botanical Garden* 72: 824–832.
- Stelkens, R., and O. Seehausen. 2009. Genetic distance between species predicts novel trait expression in their hybrids. *Evolution* 63: 884–897.
- Szymura, J. M., and N. H. Barton. 1991. The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: Comparisons between transects and between loci. *Evolution* 45: 237–261.
- Thompson, J. D., and R. Lumaret. 1992. The evolutionary dynamics of polyploid plants: Origins, establishment and persistence. *Trends in Ecology & Evolution* 7: 302–307.
- Thompson, K. A., M. Urquhart-Cronish, K. D. Whitney, L. H. Rieseberg, and D. Schluter. 2021. Patterns, predictors, and consequences of dominance in hybrids. *The American Naturalist* 197: E72–E88.
- Van de Peer, Y., T.-L. Ashman, P. S. Soltis, and D. E. Soltis. 2021. Polyploidy: An evolutionary and ecological force in stressful times. *The Plant Cell* 33: 11–26.
- Wang, D., X. Xu, H. Zhang, Z. Xi, R. J. Abbott, J. Fu, and J. Liu. 2022. Abiotic niche divergence of hybrid species from their progenitors. *The American Naturalist* 200: 634–645.
- Wang, J., J. Qin, P. Sun, X. Ma, J. Yu, Y. Li, S. Sun, et al. 2019. Polyploidy index and its implications for the evolution of polyploids. *Frontiers in Genetics* 10: e807.
- Warren, D. L., R. E. Glor, and M. Turelli. 2008. Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution* 62: 2868–2883.
- Warton, D. I., R. A. Duursma, D. S. Falster, and S. Taskinen. 2012. smatr 3—An R package for estimation and inference about allometric lines. *Methods in Ecology and Evolution* 3: 257–259.
- Whitney, K. D., R. A. Randell, and L. H. Rieseberg. 2006. Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. *The American Naturalist* 167: 794–807.
- Whitney, K. D., R. A. Randell, and L. H. Rieseberg. 2010. Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. *The New Phytologist* 187: 230–239.
- Wickham, H. 2016. ggplot2: Elegant graphics for data analysis. Springer, New York, New York, USA.
- Winge, O. 1917. The chromosome. Their numbers and general importance. *Comptes Rendus des Travaux du Laboratoire Carlsberg* 13: 131–175.
- Winge, Ö. 1932. On the origin of constant species-hybrids. *Svensk Botanisk Tidskrift* 26: 107–122.
- Zhang, W.-W., J. Song, M. Wang, Y.-Y. Liu, N. Li, Y.-J. Zhang, N. M. Holbrook, and G.-Y. Hao. 2017. Divergences in hydraulic architecture form an important basis for niche differentiation between diploid and polyploid *Betula* species in NE China. *Tree Physiology* 37: 604–616.
- Zizka, A., D. Silvestro, T. Andermann, J. Azevedo, C. Duarte Ritter, D. Edler, H. Farooq, et al. 2019. CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution* 10: 744–751.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. Sunflower trait data.

Appendix S2. Sunflower trait divergence index values.

Appendix S3. Sunflower evolution slope and delta divergence index values.

Appendix S4. R code for sunflower data analysis.

Appendix S5. *Tragopogon* trait data.

Appendix S6. *Tragopogon* trait divergence index values.

Appendix S7. R code for *Tragopogon* data used to create Figure 6.

Appendix S8. Inputs and outputs of PCA analyses.

Appendix S9. R code for PCA divergence index values and Figure 7.

Appendix S10. Description of simulations testing for patterns of overlap.

Appendix S11. R code for overlap simulations.

Video S1. A video illustrating how a hybrid or allopolyploid's biology (PP) would be characterized in DI space for a given set of parental taxa (D_1 and D_2). The x -axis represents a centroid divergence from progenitor centroids (DI_{Centroid}). The y -axis represents a breadth divergence from progenitor breadths (DI_{Breadth}). See Equations 2 and 3 in the main text.

How to cite this article: Krieg, C. P. 2024. A unified framework to investigate and interpret hybrid and allopolyploid biodiversity across biological scales. *Applications in Plant Sciences* 12(4): e11612. <https://doi.org/10.1002/aps3.11612>