INTRODUCTION

As commonly used in the English language, the term "trait" is synonymous with characteristic, attribute, or feature. In ecological parlance, a trait has been defined as any morphological, physiological, behavioural, or phenological heritable feature that can be measured or categorized at the individual level (Garnier, Navas, & Grigulis, 2016; McGill, Enquist, Weiher, & Westoby, 2006; Violle et al., 2007). Species traits, rather than taxonomic species, are increasingly acknowledged as providing new opportunities to enhance our understanding of ecological patterns and processes operating in nature (Gross et al., 2017; McGill et al., 2006). Traits have become the central component of the growing area of functional or trait-based ecology, where "functional traits" relate to the performance (growth
rate, survival, reproduction) of an organism and/or its contribution to ecological processes (Cadotte, Carscadden, & Miroetchick, 2011; Violle et al., 2007). Trait-based ecology has also gained popularity because it facilitates generalization across geographies with few species in common. For example, traits help explain patterns in species abundance (Connolly et al., 2014), explore species assemblages responses to natural and anthropogenic impacts (Mouillot, Graham, Villéger, Mason, & Bellwood, 2013), elucidate morphological variability within phylogenies (Borstein, Fordyce, O’Meara, Wainwright, & McGee, 2019), predict the outcome of species interactions (Kunstler et al., 2016), estimate likelihoods of species extinction (Luiz, Woods, Madin, & Madin, 2016) and invasion (Liu, Comte, & Olden, 2017), forecast changes in species distributions caused by climate change (Estrada, Morales-Castilla, Caplat, & Early, 2016) and, ultimately, can provide inference relating to key aspects of ecosystem function (Moore & Olden, 2017).

Fish ecologists are at the forefront of trait-based ecology (Olden et al., 2010); seconded only by plant ecologists in terms of the number of research papers published per year (Villéger, Brosse, Mouchet, Mouillot, & Vanni, 2017). The uptake of trait-based approaches by fish ecologists likely reflects the high diversity of fish forms and functions they provide to aquatic ecosystems. In fact, trait-based approaches were applied to fish ecology long before the modern definition of “functional trait” became mainstream. Many studies, particularly those published prior to 2000, used attributes that are now referred to as functional traits, but were not framed in this terminological context at the time.

Despite its widespread use, there is little consensus for an unambiguous definition of “functional trait”. For example, in coral reef ecology, a research area that focus heavily on ecosystem processes and how they are influenced by resident fish assemblages, it has been proposed that a trait should not be considered functional if it is not demonstrably linked to the movement or storage of energy or material (Bellwood, Streit, Brandl, & Tebbett, 2019). This restrictive flux-based definition makes functional traits potentially appropriate indicators of the status or resilience of ecosystems. However, a much broader range of questions has been addressed under the auspices of trait-based fish ecology, encompassing different levels of functional organization and spanning basic to applied science. To answer many of these questions, a more inclusive and feature-based definition that links the functional effects of traits to an organism’s performance and fitness is needed (McGill et al., 2006; Villéger et al., 2017; Violle et al., 2007). For example, traits have been used to model and explore the ways in which fish allocate energy for reproduction (e.g., triangular life-history model; Winemiller, 1989; Winemiller & Rose, 1992); broad-scale patterns in assemblage structure (e.g., trait-based community assembly; Fitzgerald, Winemiller, Pérez, & Sousa, 2017; Ingram & Shurin, 2009); and more recently, the development of functional diversity metrics (e.g., functional richness, evenness, divergence) and the concept of the multivariate functional space based on combinations of traits (Mouchet, Villéger, Mason, & Mouillot, 2010; Villéger, Mason, & Mouillot, 2008). The latter approaches have been proven valuable in addressing important ecological questions including: disentangling the assembly rules that shape fish assemblages (Giam & Olden, 2016; Leitão et al., 2017; Peres-Neto, 2004); understanding the effects of natural and anthropogenic disturbance on communities (Kelley, Grierson, Collin, & Davies, 2018; Mouillot et al., 2013; Rodrigues-Filho, Leitão, Zuanon, Sánchez-Botero, & Baccaro, 2018); assessing the impacts of species invasions (Olden, Poff, & Bestgen, 2006; Toussaint et al., 2018) and advancing the study of functional biogeography (Mouillot et al., 2014).

As for other scientific disciplines, breakthroughs in ecology often emerge as a result of successful connections forged between seemingly disparate areas of knowledge (Schilling & Green, 2011; Uzzi, Mukherjee, Stringer, & Jones, 2013). The ability to navigate the sea of knowledge and understand emerging ideas is a critical skill for researchers aiming at high-impact outcomes, as is the capacity to synthesize this information to generate novel concepts and methods. However, scientists tend to be siloed by the traditions of their research field, usually building incrementally upon cumulative knowledge within the bounds of their area of expertise and often resorting to lower-risk studies that re-visit previous research themes within different geographic or taxonomic contexts (McGill, 2015; Pyšek et al., 2008). Whilst such studies are important, or even essential, for deep and specialized (vertical) understanding and to fuel meta-analyses, they contribute little to fostering the cross-fertilization of ideas among disparate research areas (horizontal approach). To identify knowledge gaps and better guide the future research agenda in trait-based fish ecology, a rigorous retrospective examination of the literature is needed.

Recent advances in automated text analysis provide new opportunities to gather and scrutinize huge collections of documents, offering a unique and previously inaccessible “bird’s eye” view of
a research field. A potentially useful development has been the growth of statistical methods for performing text-mining to reveal patterns in the literature. Several of these approaches identify combinations of words within articles to help elucidate the key ideas discussed within a collection of documents (also known as a corpus). Consequently, text analysis has the potential to generate conceptual insights traditionally available only through narrative review, but with greater speed and quantitative rigour.

Here, we seek to reveal latent research topics within trait-based fish ecology by assessing the published literature using topic modelling. Topic modelling is a machine-learning method that uncovers hidden thematic structures from a corpus by identifying a set of interpretable topics that can be viewed as groups of co-occurring words that are associated with a single topic or theme (Griffiths & Steyvers, 2004). Topic modelling is commonly used in political, economic and literature sciences. However, lately this approach has shown considerable promise in ecology and conservation (Chandler, Steuckardt, Mathet, Diwersy, & Gimenez, 2018; Greenville, Dickman, & Wardle, 2017; McCallen et al., 2019; Westgate, Barton, Pierson, & Lindenmayer, 2015), and in the aquatic sciences (Syed, Borit, & Spruit, 2018). As publications continue to accumulate over time at rates difficult for researchers to embrace, the proportion of available literature covered in each new study is substantially reduced. Such reductions result in smaller representations of the entire literature, potentially hindering opportunities for the recombination of ideas among a broader array of topics within a field of research. The purpose of this study was to characterize the topic landscape of the trait-based ecology of marine and freshwater fishes, explore the inter-connections among those topics and identify research gaps and emerging avenues of investigation for the future.

2 | METHODS

2.1 | Data

The term “functional trait” is now well established in the ecological literature (McGill, 2015; McGill et al., 2006; Violle et al., 2007), yet studies analysing fish “life-history”, “ecomorphological” and “morphometric” traits that are equivalent to functional traits have been common since the 1970’s. To perform a thorough and inclusive review of the fish trait-based literature, we searched the Clarivate Analytics Web of Science titles containing the following word combination: (trait* OR functional OR life histor* OR ecomorphol* OR morphometr*) AND fish*. We collected all research papers (hereafter “articles”) from 1970 up to October 28 2018, resulting in a return of 1,954 articles. Because the words “trait” and “functional” may also represent meanings that are out of the scope of ecology (e.g., genomic traits, functional characterization of drugs), we evaluated the title and abstract of each article to ensure that only ecological trait-based studies conducted at the species-level are included in our analysis. We also eliminated articles that used the term life-history as a synonym for one specific life stage (e.g., early life-history) and did not analyse individual level traits. A total of 751 articles met the above criteria and were considered further (Table S1).

We used text from title, keywords and abstract to form the article content, and every article was tokenized (i.e., the process of obtaining individual words—also known as unigrams—from sentences). We removed meaningless terms including stop words (for example: the, or, and, which), numbers and punctuation. Minimum word length was set to three characters. The remaining words were stemmed (reduced to their base or root form, for example diverging and divergent become diverg and tested for bi-grams and tri-grams. Bi-grams are pairs of words that retain a semantic information that is lost if the words are analysed separately. For example, consider the following two sentences “The white person lives in the house” and “The person lives in the White House”. They contain the exact same words but have completely different meaning. It is common practice in text-mining analysis to connect these words with a bottom line (a process called n-gramming), therefore, making the bi-gram “white house” a unique and relevant term in political science’s text analyses. Tri-grams follow the same logic for three-word terms. We processed an automated n-gramming in our data by searching for words that occurred together more than expected by chance under a significance value of p < .05 and visually inspected the relevance and matched the n-grams with the same meaning. We then removed very rare (n < 15) and very common terms (n > 2,000) that provide little information content (Westgate et al., 2015), resulting in 901,829 entries of 1,273 unique terms.

2.2 | Analyses

2.2.1 | Topic modelling

A topic in topic models is characterized by a set of co-occurring words that bring insight into the nature of a corpus. As its name suggests, “topic modelling” can be conceptualized as a way of describing the content of different articles in a corpus (Murakami, Thompson, Hunston, & Vajn, 2017). Importantly, the categories identified by topic modelling emerge from the methodology and the corpus rather than being predetermined by an evaluator. We used latent Dirichlet allocation (LDA) modelling to identify common topics reported in our dataset. LDA identifies sets of co-occurring words that are more frequently presented within the same linguistic context than expected by chance alone. These co-occurring words tend to purport similar meaning and refer to a similar subject, thus allowing topics to be defined. The LDA model follows the assumption that articles exhibit multiple topics in mixing proportions, thus capturing the heterogeneity of the research topics within scientific publications. In statistics, this is referred to as a mixed-membership model (Erosheva, Fienberg, & Lafferty, 2004). For example, an article might be 50% about the topic “A”, 30% about the topic “B” and 20% about the topic “C”. An article is usually referred to be “about” the dominant topic (topic with the highest proportion), but this can be misleading when two or more topics occur in the same article in similar proportions or the difference between the top topics in an article is small.
Outcomes from the LDA include a list of the most common words and their topic probabilities for each article, however the model does not provide a name for each identified topic. To name each topic, we inspected the 20 most unique words (thus most representative) from each topic. When necessary, we also inspected the articles in which each topic was dominant. Due to the unsupervised nature of LDA, the number of topics is either known a priori, or chosen based on some metrics. Using the R package ‘ldatuning’ (Murzintcev, 2014), we created 50 different LDA models by varying the K-parameter from 1 to 50. The number of topics in our LDA model was selected using the optimization method proposed by Deveaud, SanJuan, and Bellot (2014). The final LDA “best” model was fitted using the R package “topicmodels” (Hornik & Grün, 2011).

2.2.2 | Topic similarity

Clusters of similar topics must first be identified before interpreting the meaning of each topic (Westgate et al., 2015). The LDA model produces a matrix of the weight (i.e., probability of occurrence) of each word within each topic that can be summarized according to an association (dissimilarity) metric and subjected to multivariate ordination analysis. This is analogous to treating words as species and topics as sites in a traditional community ecology analysis. To investigate topic similarity, we calculated the Bray–Curtis distance between each pair of topics using a matrix of the weight of each word within each topic (“word” distance matrix). We visualized the distance matrix via non-metric multidimensional scaling (NMDS).

2.2.3 | Topic popularity

To determine trends in topic popularity, the mean topic proportions for each topic were aggregated into consecutive 5-year bins. The exception to this was for the 1970–1979 decade, which was treated as a single group due to the relatively small number of papers published in this period and thereby avoids a disproportional influence of a few articles at the beginning of our sampling period. In addition, the last 4 years (2015–2018) were treated as single group. We inferred trends in topic prevalence over time by averaging topic prevalence change between two consecutive year groups. Topics with positive and negative averages in prevalence over time were considered hot topics (growing popularity) and cold topics (decreasing popularity), respectively (following Griffiths & Steyvers, 2004; McCallen et al., 2019).

2.2.4 | Topic specificity/generality

Mixed-membership is a defining characteristic of the LDA modeling; referring to the fact that an article is not assigned to a single topic, but to a combination of topics with differing proportions. Topics can then vary within a gradient from general to specific. Some topics may be general and reflect broad ideas common to many documents within the corpus, whilst others may be more specific. For example, topics that more often demonstrate a higher proportion than other topics in documents which they occur are considered more specific. Conversely, topics that more often co-occur with other topics in similar proportions may reflect broad themes common to many articles within the corpus and are considered general (Westgate et al., 2015). We used the distribution of topic weights (i.e., proportional probability of occurrence) within documents to assess the generality vs. specificity of topics. For each document, we selected the topic that received the highest weight. We then calculated the mean weight of a topic when it was selected and the mean weight of a topic when it was not selected across all documents. Plots of these values against each other for all topics were then used to compare the generality and specificity of different topics (Westgate et al., 2015).

2.2.5 | Research gap analysis

Research gaps were identified as pairs of topics that were typically separated within the corpus, both in terms of their thematic content and the articles in which they appear (Westgate et al., 2015). In addition to “word weight matrix” (see topic similarity), the LDA also produces a matrix of the weight of each topic within each article. We transposed this matrix so topics were rows and article weights were columns and then calculated the Bray–Curtis distance between each pair of topics using the matrix of topics and article weights (“article” distance matrix). We calculated the product of the “word” and “article” distance matrices (“gap” distance matrix; Westgate et al., 2015) to compare the difference in topics based on words vs. articles. We investigated research gaps between topics by plotting the dissimilarity metric from the “gap” distance matrix. The greater the metric, the larger the gap between topics (topics that both contain different sets of words and topics that rarely co-occur in the same article).

3 | RESULTS

The model selection suggested an optimal number of 16 topics for our corpus. The 20 most probable words, together with the label for each uncovered latent topic, are shown in Table 1. Some words were associated to more than one topic, resulting in topics being more or less associated to each other (Table S2 shows an extended list of words per topic). The 16 topics encompassing multiple subdisciplines of ecology that ranged from classical (population ecology, life-history) to contemporary (climate change, community ecology) and from local (biotic interactions) to broad scales (ecosystem processes). Across the study period, the most topics that dominate the largest number of articles (70 each) were population ecology and community ecology, followed by fisheries (66), ecomorphology (62), freshwater habitat and hydrogeomorphology (61) and life-history strategies (60). A visual representation of the topics, their proportions within the complete corpus and their representation in articles concerning marine and freshwater fish species are shown in Figure 1.
TABLE 1 Uncovered topics from 751 research articles about trait-based ecology of fishes published during the period 1970–2018 identified from Latent Dirichlet Allocation Modelling (LDA)

<table>
<thead>
<tr>
<th>Topic number</th>
<th>Topic name</th>
<th>Top topic words (stemmed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fisheries</td>
<td>fisheri, exploit, stock, biomass, harvest, catch, yield, target, fisheri_pressur, commerci, fisheri_manag, sustain, effort, protect_area, potenti, estim, assess, bear, by, catch, limit</td>
</tr>
<tr>
<td>2</td>
<td>Ecomorphology</td>
<td>morpholog_trait, rich, ecomorpholog, trophic, diet, phylogenet, resourc, evolutionari, form, overlap, morphometr, partit, diversif, bodi_shape, dimens, eye, multivari, mouth, phylogeni, space, converg</td>
</tr>
<tr>
<td>3</td>
<td>Invasion ecology and extinction</td>
<td>rang, nativ, mediterranean_sea, invas, success, establish, european, extinct, red_list, introduc, endem, alien, threaten, status, endang, exot, introduc, invalid, lake, genet_divers</td>
</tr>
<tr>
<td>4</td>
<td>Population ecology</td>
<td>popul, life_histori_evolut, bodi_size, growth_rate, age_matur, mortal, adult, recruit, surviv, densiti_depend, demograph, life_stage, regul, fisheri_induc_evolut, lifetime, phenotyp_plastic, dynam, reproduct_invest, stock, genet_variat</td>
</tr>
<tr>
<td>5</td>
<td>Migration</td>
<td>estuar, migrat, coastal, otolith, connect, salin, guild, bay, migratori, diadrom, juvenil, marsh, stabl_isotop, delta, resid, amphidrom, movement, life_cycl, nursery, complex</td>
</tr>
<tr>
<td>6</td>
<td>Ecosystem processes</td>
<td>ecosystem, vulner, resili, herboriv, coral_reef, phase_shift, global, conserv, limit, loss, physiolog, key, nutrient, aggreg, knowledge, process, metabol, world, applic, future</td>
</tr>
<tr>
<td>7</td>
<td>Life-history strategies</td>
<td>reproduct, life_histori_strategi, season, spawn, fecund, egg_size, energi, parent_care, trade_off, offspr, alloc, opportunist, equilibrium, continuum, matern, cost, spawner, gonad, bodi_size, surviv</td>
</tr>
<tr>
<td>8</td>
<td>Freshwater habitat and hydrogeomorphology</td>
<td>river, habitat, stream, flow, basin, dam, hydrolog, fish_assemblag, restor, floodplain, downstream, forest, channel, upstream, flood, pool, substrat, headwat, long_term, mesohabitat</td>
</tr>
<tr>
<td>9</td>
<td>Climate change</td>
<td>respons, sensit, increase, impact, declin, consequ, climat, chang, trend, intens, project, stressor, ecosystem, warm, year, magnitud, refug, pollut, anthropogen, arctic, degrad, contamin, disturb</td>
</tr>
<tr>
<td>10</td>
<td>Biotic interactions</td>
<td>variat, contrast, individ, environ, diverg, phenotyp, compet, interspecif, adapt, latitudin, constrain, frequent, intraspecif, sympat, size, pattern, latitud, north, thermal, temper</td>
</tr>
<tr>
<td>11</td>
<td>Analytical approaches</td>
<td>predict, analysi, relationship, approach, correl, data, model, test, cluster, appli, framework, index, theoret, tool, distribut, varianc, random, regress, statist, matrix</td>
</tr>
<tr>
<td>12</td>
<td>Landscape ecology</td>
<td>environment, spatial_variat, region, local, spatial_scale, gradient, tempor, geograph, site, biotic, landscap, process, abiot, driver, area, speci_pool, biogeograph, stream_fish, catchment, continent</td>
</tr>
<tr>
<td>13</td>
<td>Marine ecology</td>
<td>larval, water_temperature, ocean, marin, coral_reef_fish, dispers, distribut, island, pelag, geogr_rang, settlement, deep_water, sea, distanc, benthic, demers, shelf, barrier, coral_reef, genet</td>
</tr>
<tr>
<td>14</td>
<td>Community ecology</td>
<td>function_trait, function_divers, fish_assemblag, fish_community, environment_filter, taxonom, redund, beta_divers, turnov, non_nativ, environment_gradient, null_model, phylogenet_divers, metric, disturb, conserv, guild, function_rich, function_structur, speci_rich</td>
</tr>
<tr>
<td>15</td>
<td>Behavioural ecology</td>
<td>male, femal, behavio, condit, mate, sexual_select, colour, parasit, sex, social, fit, larger, cichlid, exhibit, nest, tactic, genet, femal_prefer, sex_ratio, sperm</td>
</tr>
<tr>
<td>16</td>
<td>Trophic ecology</td>
<td>predat, feed, prey, interact, forag, swim, food, pector_fin, behavio, morpholog, food_web, jaw, speed, caudal, trophic, consumpt, locomot, cascad, captur, ontogeni</td>
</tr>
</tbody>
</table>

Note: Each topic displays the 20 words with highest probability of occurrence. The topics were assigned a name that best describe the semantics of the top words. An extended version of this table with the top 70 words in each topic is provided in the supplemental material (Table S2).

The prevalence of topics (i.e., the proportional contribution of each topic in the document corpus) represented in the literature varied considerably over the last half century (Figure 2). Throughout the 1970s and 1980s, the most popular topics included life-history strategies, trophic ecology, population ecology, biotic interactions, and freshwater habitat and hydrogeomorphology (Figure 2, Figure S2). Among them, life-history strategies, biotic interactions and trophic ecology demonstrated significant decreases in prevalence through time and are thus considered “cold topics” of current interest in the literature. On the other end of the popularity spectrum, community ecology became the most dominant topic for the time period after 2010 (Figure 2), and together with the ecosystem processes, and analytical approaches are three “hot topics” in the literature, demonstrating significant increase in topic prevalence through time. Cold topics were not topics that necessarily lacked prevalence in the corpus but those that showed a strong decrease in popularity over time. For example, life-history strategies and trophic ecology showed strong declines in prevalence, but they are still relatively well represented in the corpus during the last decade. Similarly, hot topics do not necessarily have high weight prevalence overall (Figure S3). The community ecology topic had negligible representation in the early years of this analysis, but then increased to ~5% prevalence in the period after 2005, which coincides with the publication of the first papers about fish that used the term “functional trait” (Figure 2b). The more specific a topic, the more likely it was to be the sole focus (and hence the highest weighted topic) of a document. We
found that behaviour ecology and trophic ecology were highly specific topics (Figure 3). By contrast, analytical approaches, landscape ecology, biotic interactions, ecosystem processes and climate change were general (Figure 3), which indicated that these topics were broader and therefore often discussed in association with other topics.

Analysis of research gaps showed that some topics that have changed most in prevalence (e.g., community ecology, trophic ecology; Figure 2) and topics that are very specific (behavioural ecology, trophic ecology; Figure 3) have the largest gaps on average across all topics (Figure 4). Of these, community ecology and trophic ecology (the fastest growing and fastest decreasing topics, respectively) and behavioural ecology (the most specific) ranked in the top three in terms of separation from the remaining topics. By contrast, the most general topics representing broader scientific concepts (e.g., analytical approaches, biotic interactions, marine ecology) had higher overall connectivity with other topics. The largest research gaps found, far exceeding all other topics, was between community ecology and the population ecology, behavioural ecology and life-history strategies topics.

4 | DISCUSSION

We tracked the evolution of trait-based research in the field of fish ecology applying text mining to peer-reviewed articles published over the past half century. By mapping the topic landscape of the literature, we found that articles are structured into 16 latent topics that vary in frequency, popularity and specialization. Two of those topics—population ecology and community ecology—stand out as very influential. They individually dominated topic weight in more papers than any other topic.

Life-history traits of fishes have been applied to studies in population ecology since the 1970’s, which then spawned the more theoretical topic of life-history strategies. The connection among these topics is evidenced by the word-use similarity they shared (Figure 1). The community ecology topic, on the other hand, was poorly investigated in the earlier years of our dataset but grew in prevalence during the past decade at a faster rate than any other to become one of the dominant topics in the field. Interestingly,
**FIGURE 2** Change in prevalence in the literature over the study period for each of the 16 topics. (a) Each dot represents the average prevalence change among consecutive year groups (shown in d); bars represent standard errors. See Figure S2 for topic/year-group values. Plots at the right side of the panel tracks the temporal changes in topic prevalence of (b) hot, (c) neutral and (d) cold topics. Figure appears in colour in the online version only [Colour figure can be viewed at wileyonlinelibrary.com]

**FIGURE 3** Topic generality/specificity. Topics in the top left-hand corner are specific (more likely to be the main topic present within an article), whilst topics in the bottom right are general (broad topics common to many articles within the corpus).
the community ecology topic shared high word-use similarity with the trophic ecology topic. We interpret this connection as the result of many studies on the community ecology topic being strongly grounded on traits related to resource acquisition, specifically to morphological traits related to food detection and capture (e.g., eye size, gape size, mouth position) and ecological traits based on diet or food items categories (e.g., detritivores, herbivores, predators; Villéger et al., 2017). Taken together, these five topics form two cohesive groups that dominate topic weight in 42% of the articles in our corpus.

Scientific interest in topics waxes and wanes over time, although the reasons are generally poorly known (Griffiths & Steyvers, 2004). To gain insight into the dynamics of a research field, we must understand the dynamics of its scientific topics (Mane & Börner, 2004). In an anecdotal note, McGill (2015) critically reflected that after the publication of landmark papers that formalized definitions for functional traits (i.e., McGill et al., 2006; Violle et al., 2007), functional ecology research using traits became excessively popular in community ecology, in part, because researchers perceived it as a useful context to broaden the relevance of their work. Our data for fish trait-based research support McGill’s argument regarding the timing of uptake of this terminology (Figure 2); however, our analyses do not shed further light on the motivations of the community ecologists who adopted functional traits as terminology during this time. Some changes in topic prevalence may be due to shifts in methods and theoretical developments rather than a change in scientific interest. For example, the trophic ecology topic, which bears a strong word-use similarity to community ecology, declined in prevalence over the same period that community ecology increased. Apparently, as statistical and computer power developed, studies in trophic ecology may have naturally transitioned from early stomach content and ecomorphological studies analysing single species and traits to the multispecies, multi-trait multivariate functional space approaches that characterize recent studies in the community ecology topic.

Perhaps the most insightful aspect of topic modelling is the ability to objectively assess research gaps among topics in a quantitative manner. Such gaps can be identified as pairs of topics that are unusually separate within the corpus, both in terms of their thematic content and the articles in which they appear. Scientific innovation is unlikely to flourish from research performed within narrow
boundaries, and the most creative ideas often arise via unusual combinations of subjects that spark new insights (Schilling & Green, 2011; Uzzi et al., 2013). A practical outcome of our analysis is the identification of knowledge gaps among trait-based approaches applied to fish ecology that will likely result in innovative advances and therefore can be used to guide researchers as to where they can look for novel or emerging topics, as well as research gaps where research effort might best be directed.

Our results suggest that there were varying gaps in information among topics. Community ecology showed the largest average degree of separation from the other 15 topics, and the largest pairwise gaps we detected were between community ecology and population ecology (95% dissimilarity), behavioural ecology (92%) and life-history strategies (83%; Figure 4). The lack of connectivity between community ecology and other topics can be explained by its recent association to functional traits and that, given enough time, it will expand and begin to associate more with other topics. Another mechanism that underlies these gaps around the community ecology topic is the "bandwagon effect" (Didham, Leather, & Basset, 2016; McGill, 2015), or the tendency of many "me-too" studies (i.e., repeats of research themes already examined in other regions or contexts). In the case of fish trait-based research, the bandwagon effect can be reinforced by limited availability of trait datasets, which encourage researchers to analyse the same trait types repeatedly. For example, during the writing of this review, it became apparent that community ecology studies have disproportionately used large scale occurrence-only datasets and focused on morphological traits derived from body measurements and trophic traits such as diet items or trophic position which explains the large gaps between the community ecology and the population ecology and life-history strategies topics, as the latter two are mostly based on reproductive and physiological related traits (e.g., age of maturation, clutch size, growth date, life span). The potential for ecologists' background to influence their trait choice is apparent by the partitioning of topics into distinct overarching themes, namely: study system (marine ecology, freshwater habitat), biological mechanisms (biotic interactions, behavioural ecology, life-history, migration), methodology (ecomorphology, analytical approaches, landscape ecology), drivers (fisheries, climate change, ecosystem processes, invasion ecology) and scale of organization (population, community). We suggest that future community level studies integrating functional diversity metrics and the concept of multivariate functional space with reproductive traits that may influence energetic allocation and population growth across species abundances will generate fruitful advances. This will not be accomplished without challenges. Information on life-history traits are more difficult to collect than traits based on external measurements or by underwater observations. In addition, simultaneously collecting accurate quantitative (biomass or abundance) and functional trait data are very demanding. There are, however, encouraging examples showing that collecting life-history traits for large assemblages on a continental scale is an attainable endeavour (Giam & Olden, 2016; Sternberg & Kennard, 2014; Teichert et al., 2017). Collaborations among researchers with different backgrounds and skills targeting these research gaps will be instrumental to advance our understanding of fish ecology.

A final consideration is the potential for the topics identified by research gap analysis to identify areas of strong specialization. Within the fish trait-based literature, behavioural ecology and trophic ecology showed a large degree of separation from the other topics identified. It is possible that this gap emerged because these topics are particularly specialized in terms of the traits examined and the methods used to collect data relating to fish behaviour. Such a finding would suggest that there are opportunities for novel and innovative research that integrates behavioural ecology with other topics in fish ecology research. We caution, however, that deciding which of the research gaps identified represent fruitful directions for future research is not always straightforward. Some combinations of topics we identified (Figure 4) may have been rare because researchers' ability or desire to combine insights from these distinct fields is limited (e.g., due to logistical issues or a perceived lack of synergy), rather than because they were overlooked. Nonetheless, we believe that quantitative assessment of literature trends is a useful approach for identifying key knowledge gaps and potential directions for future research in the field of fish ecology.

ACKNOWLEDGEMENTS
This research was funded by the Australian Government through an Australian Research Council grant (LP150100388).

CONFLICT OF INTEREST
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
Appendix S1 contains a table with information on all articles analysed in this study.

ORCID
Osmar J. Luiz https://orcid.org/0000-0002-6995-6524
Julian D. Olden https://orcid.org/0000-0003-2143-1187
Alison J. King https://orcid.org/0000-0002-0104-2611

REFERENCES


**SUPPORTING INFORMATION**

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