

Original article

연관 규칙 학습과 군집분석을 활용한 멸종위기 기수갈고둥과 생태계 내 종 간 연관성 분석

임성호^{1,2} · 도윤호^{1,*}

¹(주)생태조사단 부설 두희생태연구소, ²공주대학교 생명과학과

Analyzing the Co-occurrence of Endangered Brackish-Water Snails with Other Species in Ecosystems Using Association Rule Learning and Clustering Analysis. Sung-Ho Lim^{1,2} (0000-0002-4365-9695) and Yuno Do^{1,*} (0000-0001-7910-9751) (¹Doohee Institute of Ecological Research, Korea Ecosystem Service Inc., Ansan 15426, Republic of Korea; ²Department of Biological Sciences, Kongju National University, Gongju 32588, Republic of Korea)

Abstract This study utilizes association rule learning and clustering analysis to explore the co-occurrence and relationships within ecosystems, focusing on the endangered brackish-water snail *Clithon retropictum*, classified as Class II endangered wildlife in Korea. The goal is to analyze co-occurrence patterns between brackish-water snails and other species to better understand their roles within the ecosystem. By examining co-occurrence patterns and relationships among species in large datasets, association rule learning aids in identifying significant relationships. Meanwhile, K-means and hierarchical clustering analyses are employed to assess ecological similarities and differences among species, facilitating their classification based on ecological characteristics. The findings reveal a significant level of relationship and co-occurrence between brackish-water snails and other species. This research underscores the importance of understanding these relationships for the conservation of endangered species like *C. retropictum* and for developing effective ecosystem management strategies. By emphasizing the role of a data-driven approach, this study contributes to advancing our knowledge on biodiversity conservation and ecosystem health, proposing new directions for future research in ecosystem management and conservation strategies.

Key words: brackish-water snail, association rule learning, market basket analysis, estuary, endangered species

서 론

생태계는 지구상의 모든 생명체가 상호 의존하는 복잡한 네트워크를 형성하며, 이는 생물 다양성의 보존을 통해 유지된다(Thorpe *et al.*, 2006). 생물 다양성은 생태계의 건강과 안

정성을 유지하는 데 필수적이며, 다양한 생물종들은 생태계 서비스 제공, 식량 안보, 질병 조절 그리고 기후 변화 완화와 같은 중요한 역할을 수행한다(Pedersen Zari *et al.*, 2022). 그러나 인간 활동으로 인한 서식지 파괴, 기후 변화, 오염 그리고 외래종의 침입은 전 세계적으로 생물 다양성 감소를 초래하고 있다. 이러한 상황에서 생물 다양성의 보존과 지속 가능한 관리는 인류와 자연 환경의 미래에 있어 중대한 과제로 부각되고 있다(Keil *et al.*, 2015). 특히, 특정 생태계 내에서 중요한 역할을 하는 개별 종들의 보존은 이러한 과제를 해결

Manuscript received 3 January 2024, revised 28 May 2024,
revision accepted 28 May 2024
* Corresponding author: Tel: +82-41-850-8501, Fax: +82-41-850-8501
E-mail: doy@kongju.ac.kr

하는 데 있어 핵심적인 요소가 된다(Shogren *et al.*, 1999).

이러한 상황에서 한국에서 멸종위기야생생물 II급으로 지정된 기수갈고둥(*Clithon retropictum*)이 차지하는 역할은 매우 중요하다. 기수갈고둥은 기수역과 같이 제한된 환경에서 서식하는 멸종 위기종으로, 이들의 출현 여부는 하구 생태계의 상태를 간접적으로 이해할 수 있는 정보를 제공한다(Jang *et al.*, 2021; Park *et al.*, 2022). 기수갈고둥의 생태계 내 역할이 명확히 제시되지는 않았지만 기수역은 담수와 해수가 만나는 독특한 환경으로, 특정한 생물종만이 적응하여 살아갈 수 있으며, 기수갈고둥은 이러한 환경에서 물질 순환을 촉진하고, 해양 및 담수 생태계를 연결하는 역할 등을 가질 것으로 추측할 수 있다(Cognetti and Maltagliati, 2000). 한국과 같이 대부분의 하구가 개발로 인해 닫힌 상태인 곳에서 기수갈고둥은 하구의 건강성을 판단하는 데 사용될 수 있는 생물 지표종(bioindicator species)으로 연구되어 왔다. 그러나 인간 활동에 의한 서식지 파괴로 인해 이들의 존재가 위협받고 있어, 기수갈고둥의 보존은 하구 생태계의 건강과 지속 가능성을 유지하기 위한 중요한 과제가 되었다. 따라서, 기수갈고둥과 같은 멸종 위기종의 생태학적 중요성을 이해하고, 그들의 보존을 위한 효과적인 전략을 개발하는 것이 필수적이다.

기수갈고둥을 비롯한 멸종위기야생생물을 조사할 때 몇 가지 반복해서 제시되는 어려움이 있다(Szaro, 2008; Shaw *et al.*, 2021). 멸종위기야생생물은 대체로 제한된 지역에만 서식하고, 개체 수가 적어 관찰과 채집이 까다롭다. 더욱이, 이들의 생태적 특성과 서식지 요구 사항을 파악하려면 장기간에 걸친 현장 연구가 필수적이다. 그러나 인간 활동으로 인한 서식지 파괴와 환경 변화가 빠르게 일어나면서, 이러한 전통적인 연구 방법론으로는 종의 변화를 신속하게 감지하고 대응하는 데 한계가 있다. 여러 어려움을 극복해서 획득한 자료라고 하더라도 때로 해당 생물종과 같은 공간에 서식하는 여러 생물 간의 상호작용 및 환경 변화에 대한 반응을 설명하지 못하는 경우가 많다. 전통적인 생태학적 연구 방법은 종종 환경 변수의 측정과 분석에 의존한다. 이러한 접근법은 환경 변수들이 생물종의 출현과 분포에 미치는 영향을 이해하기 위해 중요하지만, 여러 가지 한계가 있다. 첫째, 연속된 환경 변수 자료를 얻기가 어렵고, 둘째, 측정 가능한 환경 변수가 제한적이어서 특정 생물종과의 관계가 명확하게 드러나지 않을 수 있다. 예를 들어, 기후 변화, 수질 변화, 서식지 변화 등 다양한 환경 요인이 생물종의 출현에 영향을 미치지만, 이들 각각의 요인과 특정종의 출현 양상 사이의 관계를 명확하게 규명하기가 어렵다(Lindenmayer and Likens, 2010; Dale and Beyeler, 2001).

멸종위기야생생물의 조사와 연구에 있어 전통적인 방법론이 가지는 한계를 극복하는 데 연관 규칙 학습(association

rule learning)이 유용하게 적용될 수 있다(Leote *et al.*, 2020). 연관 규칙 학습은 복잡하고 다차원적인 생태 데이터를 효율적으로 분석하는 데이터 마이닝 기법으로, 대규모 데이터 세트에서 생물종 간의 연관성 및 패턴을 찾아낼 수 있다. 이 방법론은 원래 장바구니 분석(market basket analysis)에서 주로 적용되었다. 장바구니 분석은 고객의 구매 기록 분석을 통해 서로 다른 상품 간의 구매 연관성을 밝혀내는 기법으로, 특정 상품을 구매할 때 함께 구매되는 상품의 경향을 파악한다. 예를 들어, “빵을 구매하는 고객 중 60%가 버터도 함께 구매한다”는 식의 패턴을 발견할 수 있다. 이러한 정보는 상품 배치, 프로모션 전략, 재고 관리에 중요한 통찰을 제공한다. 이와 같은 원리를 생태학적 연구에 적용하면, 연관 규칙 학습을 통해 특정 서식지에서 서로 다른 종들이 어떻게 공동으로 출현하며, 이러한 공동출현이 생태계 내에서 어떤 역할을 하는지를 파악할 수 있다(Samecka-Cymerman *et al.*, 2010). 서식지에서 관찰되는 여러 종은 그 서식지의 환경적 요소에 대한 오랜 기간의 반응과 적응의 결과이다. 이러한 접근 방식은 생물종의 출현 패턴을 통해, 환경 변수의 직접적 측정 없이도 서식지 내 환경 조건과 그리고 생물종 간의 상호작용을 간접적으로 이해할 수 있게 해준다(Aronne *et al.*, 2012; Souza *et al.*, 2021).

본 연구의 목적은 연관 규칙 학습과 군집분석을 활용하여, 기수갈고둥과 공동으로 출현하는 다른 생물종을 파악하고 생태계 내 역할을 분석하는 것이다. 특히, 기수갈고둥의 출현 양상과 다른 저서성 대형무척추동물들과의 공동출현 패턴을 파악하고, 이러한 공동출현이 기수갈고둥의 서식지 선택 및 분포를 파악하는 데 도움이 되는지 이해하고자 한다. 또한, 본 연구는 새로운 접근 방식을 채택하여, 생태계 내 복잡한 종 간의 출현 현황을 정량적으로 파악하고 이를 통해 기수갈고둥과 같은 멸종위기종의 보존 전략 개발에 기여할 수 있는 근거를 마련하고자 한다.

재료 및 방법

1. 자료수집

본 연구에서 사용된 저서성 대형무척추동물의 자료는 환경부에서 수행 중인 “하구 수생태계 현황 조사 및 건강성 평가”(MOE/NIER) 연구용역의 결과물에서 추출했다. 대상지 점은 환경부 “물환경측정망 설치·운영 계획”(MOE, 2016)에 고시된 총 325개 하구 중에서 선택된 열린하구 201개 지점에 대한 수생태계 조사 결과를 이용하였으며, 2016년부터 2018년까지의 기간동안 지점 당 홍수기 전·후 2회 조사를 수행, 총 72종의 저서성 대형무척추동물을 포함하고 있다. 본

연구에서는 이 자료를 기반으로 하여 기수갈고등을 비롯한 다른 저서성 대형무척추동물의 출현 패턴을 분석하고, 연관 규칙 학습을 적용하여 이들 생물종 간의 공동출현 파악 및 서식지 조건과의 연관성을 분석했다.

2. 자료분석

데이터 전처리 단계에서, 수집된 저서성 대형무척추동물의 지점별 각 종의 개체수 자료를 불리언(boolean) 형식으로 변환하여 수치가 0보다 큰 경우 참으로, 그렇지 않은 경우 거짓으로 처리했다. 이후, 이 데이터를 트랜잭션(transaction) 형식으로 변환하여 연관 규칙 분석에 사용했다(Fig. 1).

연관 규칙 분석 과정에서는 트랜잭션 데이터를 사용하여 apriori 알고리즘을 적용했다. 이 과정에서 지지도(support)는 0.001, 신뢰도(confidence)는 0.5, 규칙의 최소 길이(minlen)는 2, 최대 길이(maxlen)는 20로 설정했다. 지지도는 규칙이 데이터 세트 내에서 얼마나 자주 발생하는지, 신뢰도는 한 항목이 발생했을 때 다른 항목도 발생할 조건부 확률을, 향상도(lift)는 규칙의 항목들이 독립적으로 발생하는 경우와 비교하여 얼마나 더 자주 함께 발생하는지를 나타낸다. 각 규칙은 이들 지표에 따라 평가되었으며, 향상도와 신뢰도에 기반해 다양한 카테고리로 분류했다.

분류 로직에는 'if-then' 구조를 사용했다. 이 구조는 규칙의 향상도와 신뢰도 값을 기준으로 규칙을 세분화하는

데 사용됐다. 향상도가 1.5보다 크면, 이 규칙은 특히 강력한 연관성을 가진 것으로 간주했다. 이러한 규칙 중에서 신뢰도가 1인 경우, 즉 항상 참인 경우는 "High lift and high confidence"로 분류되어, 매우 강력하고 신뢰할 수 있는 규칙으로 평가했다. 반면 향상도는 높지만 신뢰도가 1이 아닌 경우에는 "High lift but not high confidence"로 분류되어, 여전히 강력하지만 신뢰도가 낮은 규칙으로 평가했다. 마지막으로 향상도가 1.5 이하인 규칙은 "Low or moderate lift"로 분류되어, 상대적으로 연관성이 낮거나 중간 정도인 규칙으로 간주했다.

본 연구는 기존 군집분석 방법과 연관 규칙 학습의 효율성을 비교하고자 K-means 군집분석과 계층적 군집분석(hierarchical clustering)을 사용했다. K-means 군집분석에는 엘보우 방법(Elbow method)을 적용해 최적의 군집(cluster) 수를 결정했고, 이를 통해 데이터 포인트들을 그들의 군집 중심까지의 거리 합이 최소가 되는 방식으로 분류했다. 각 군집에서 최빈도로 출현하는 종을 기준으로 지표종을 선정하였다. 계층적 군집분석은 Ward의 방법을 사용해 종 간 거리를 바탕으로 군집을 생성하고, 이들의 계층 구조를 덴드로그램(Dendrogram)으로 시각화했다. 이 분석들을 수행하기 위해 R (version 4.2)에서 arules (Hahsler *et al.*, 2005), arulesViz (Hahsler, 2017), cluster (Maechler, 2018), stat (R Core Team, 2013) 패키지를 활용했다.

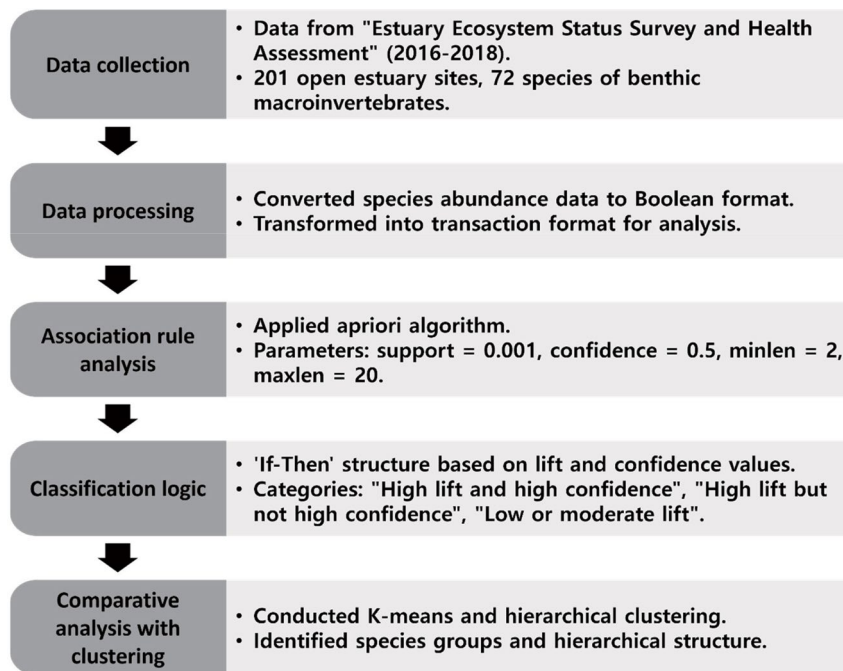


Fig. 1. Methodological steps in *Clithon retropictum* ecological analysis using association rule learning.

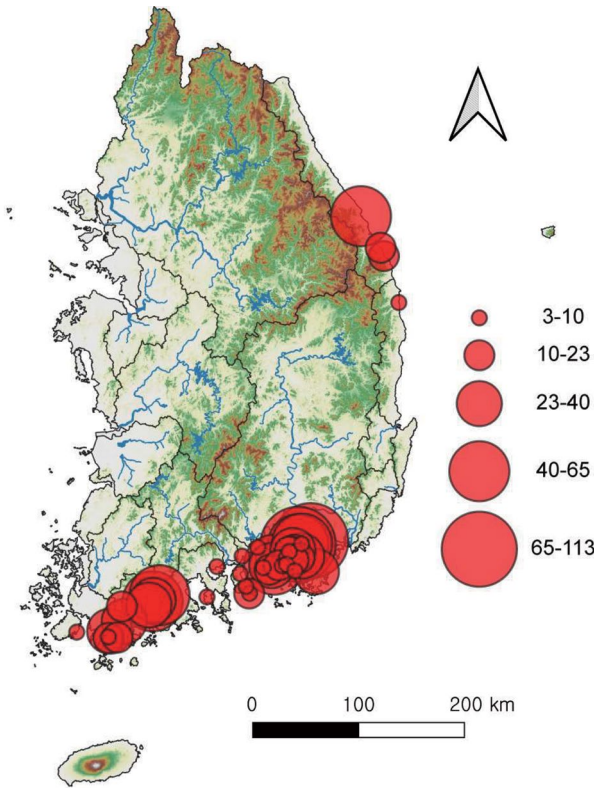


Fig. 2. Distribution of *Clithon retropictum*.

결 과

1. 기수갈고둥 분포

기수갈고둥의 분포는 주로 남해안 일대에 집중되어 있으며, 동해안의 일부 지역에서도 발견되었다(Fig. 2). 전체 201개 조사 지점 중에서 기수갈고둥이 발견된 지점은 45개로, 이는 전체 지점의 약 22.3%를 차지한다. 이들 지점에서 기수갈고둥의 평균 개체밀도는 24.2개체 m^{-2} 로 확인되었고, 표준편차는 23.23개체 m^{-2} 로 나타났다. 이는 개체밀도에 큰 변동성이 있음을 나타내며, 중앙값은 15개체 m^{-2} , 최빈값은 8개체 m^{-2} 로 측정되었다. 이 결과는 기수갈고둥의 서식지별 개체밀도 분포에 상당한 차이가 있음을 보여준다.

2. 조사 지점별 종 다양성 및 출현 빈도

하구 지점에서 해당 종이 발견된 빈도수(Frequency)는 해당 아이템이 전체 트랜잭션(지점) 중 얼마나 자주 등장하는지를 나타내며, 관찰 횟수는 해당 아이템(종)이 전체 조사 지점에서 몇 번 관찰되었는지를 나타낸다(Fig. 3a). 참갯지렁이류(Nereidae sp.)는 전체 관찰 중 69.15%에 해당하는 139회 관찰되었으며, 깔따구류(Chironomidae sp.)는 68.66%에

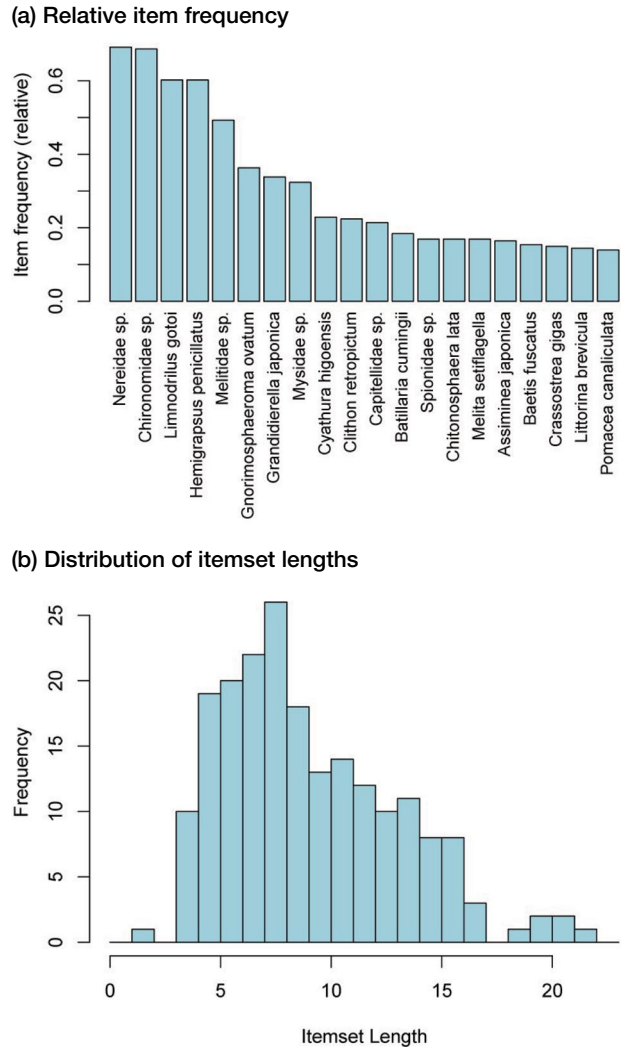


Fig. 3. Comprehensive analysis of transaction data: (a) Relative item frequencies and (b) distribution of itemset lengths.

해당하는 138회 관찰되었다. 실지렁이(*Limnodrilus gotoi*)와 풀게(*Hemigrapsus penicillatus*)는 각각 약 60.20%의 빈도로 121회 관찰되었다. 멜리타옆새우류(*Melitidae* sp.)는 49.25%의 빈도로 99회 관찰되어, 상위 20개 아이템 중 다섯 번째로 높은 빈도를 보였다. 추가로, 개펄잔벌레(*Gnorimosphaeroma ovatum*)는 73회(36.32% 빈도), 발성육질꼬리옆새우(*Grandidierella japonica*)는 68회(33.83% 빈도), 곤쟁이류(*Mysidae* sp.)는 65회(32.34% 빈도) 관찰되었다. 모래마디벌레(*Cyathura higoensis*)는 22.89%의 빈도로 46회, 기수갈고둥(*Clithon retropictum*)은 22.39%의 빈도로 45회 관찰되었다. 나머지 상위 아이템들로는 버들갯지렁이류(*Capitellidae* sp.) (43회 관찰), 땀가리(*Batillaria cumingii*) (37회 관찰), 얼굴갯지렁이류(*Spionidae* sp.) (34

회 관찰), 납작잔벌레(*Chitonosphaera lata*) (34회 관찰), 큰손멜리타옆새우(*Melita setiflagella*) (34회 관찰), 기수우렁이(*Assiminea japonica*) (33회 관찰) 등이 포함되었다. 트랜잭션 길이는 최소 2개에서 최대 22개 사이로 다양하게 분포하며, 대부분의 트랜잭션은 중간 범위의 아이템 개수를 보였다 (Fig. 3b). 가장 많이 관찰된 트랜잭션 길이는 약 6개에서 10개 사이였다. 특히, 8개와 9개 길이의 트랜잭션이 빈번하게 나타났는데, 이는 하구 지점별로 비교적 균일한 수의 종이 관찰되었음을 나타낸다. 반면, 15개 이상의 아이템을 포함하는 트랜잭션은 상대적으로 드물게 나타났다.

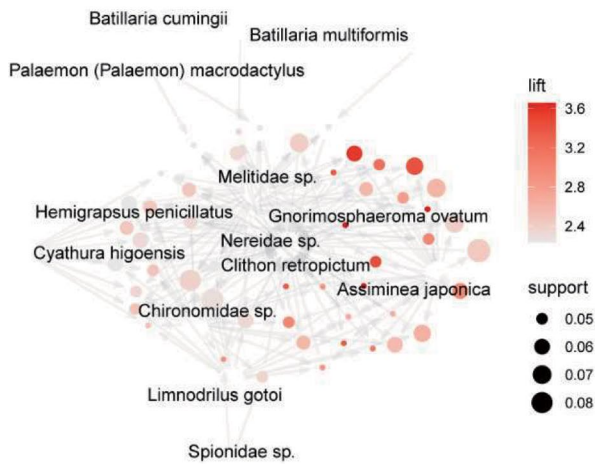
3. 기수갈고등 출현과 연관된 종들의 공동출현 분석

연관 규칙 분석 결과, 향상도 값에 따라 규칙의 개수는 크게 다양했다. 가장 높은 향상도 값인 4.47에서는 1,007,498

개의 규칙이 도출되었으며, 이는 전체 규칙 중 가장 많은 비율을 차지했다. 반면, 향상도 값이 2.23인 규칙은 35,561개로 나타났다. 지지도 값에 대해서는 0.5% 이상에서 1,028,932개의 규칙이 관찰되었고, 이는 전체 규칙의 대부분을 구성했다. 지지도 값이 증가함에 따라 규칙의 수는 감소하는 경향을 보였다. 신뢰도 분석에서는 신뢰도 값이 1인 규칙이 1,007,498개로, 전체 규칙 중 상당한 비율을 차지했다.

연관 규칙 분석 결과에서 지지도에 기반한 상위 50개 규칙의 분석을 통해, 기수갈고등의 출현과 연관된 다양한 종들을 확인했다 (Table 1, Fig. 4a). 이 중 가장 높은 지지도를 보인 규칙은 기수우렁이(*A. japonica*)가 나타날 때 기수갈고등이 출현하는 경우로, 이 규칙은 전체 관찰 중 약 8.96%의 지지도와 54.55%의 신뢰도를 보여주었다. 이어서 참갯지렁이류(*Nereidae* sp.), 실지렁이(*L. gotoi*), 멜리타옆새우

(a) Top 50 association rules by support



(b) Top 50 association rules by lift

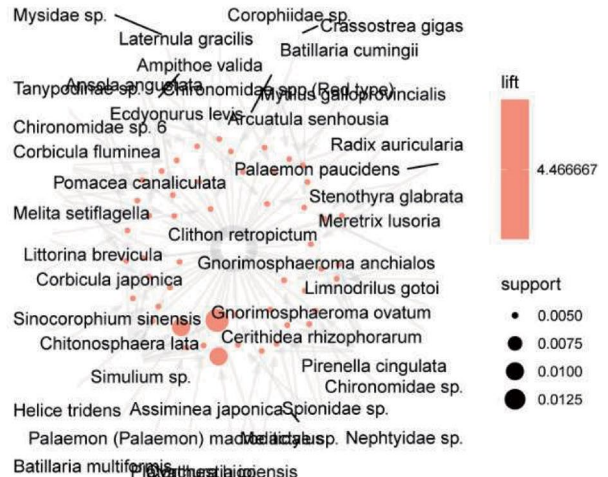


Fig. 4. Visualization of top 50 association rules by support (a) and lift (b).

Table 1. Prominent rules influencing *Clithon retropictum* presence by support.

#	LHS (Antecedent)	RHS (Consequent)	Support	Confidence	Coverage	Lift	Count
1	{ <i>Assiminea japonica</i> }	{ <i>Clithon retropictum</i> }	0.090	0.545	0.164	2.436	18
2	{ <i>Nereidae</i> sp., <i>Limnodrilus gotoi</i> , <i>Melitidae</i> sp., <i>Chironomidae</i> sp.}	{ <i>Clithon retropictum</i> }	0.085	0.515	0.164	2.301	17
3	{ <i>Nereidae</i> sp., <i>Limnodrilus gotoi</i> , <i>Melitidae</i> sp., <i>Hemigrapsus penicillatus</i> , <i>Chironomidae</i> sp.}	{ <i>Clithon retropictum</i> }	0.080	0.533	0.149	2.382	16
4	{ <i>Assiminea japonica</i> , <i>Hemigrapsus penicillatus</i> }	{ <i>Clithon retropictum</i> }	0.070	0.583	0.119	2.606	14
5	{ <i>Assiminea japonica</i> , <i>Nereidae</i> sp.}	{ <i>Clithon retropictum</i> }	0.070	0.538	0.129	2.405	14

LHS (Left-Hand Side): This column shows combinations of conditions or entities, RHS (Right-Hand Side): This is the outcome or effect of the conditions in the LHS, Support: Indicates the frequency of the rule in the dataset, Confidence: Shows the probability of the RHS occurring when the LHS is true, Coverage: The proportion of the dataset that contains the LHS, Lift: A measure of the strength of the association between the LHS and RHS. A lift greater than 1 indicates a strong association, Count: Number of times the rule is observed in the dataset.

Table 2. Prominent rules influencing *Clithon retropictum* presence by support by lift.

#	LHS (Antecedent)	RHS (Consequent)	Support	Confidence	Coverage	Lift	Count
1	{ <i>Stenothyra glabrata</i> , <i>Pirenella cingulata</i> }	{ <i>Clithon retropictum</i> }	0.005	1.000	0.005	4.467	1
2	{ <i>Pirenella cingulata</i> , Nephthyidae sp.}	{ <i>Clithon retropictum</i> }	0.005	1.000	0.005	4.467	1
3	{ <i>Cerithidea rhizophorarum</i> , <i>Pirenella cingulata</i> }	{ <i>Clithon retropictum</i> }	0.005	1.000	0.005	4.467	1
4	{ <i>Assiminea japonica</i> , <i>Pirenella cingulata</i> }	{ <i>Clithon retropictum</i> }	0.005	1.000	0.005	4.467	1
5	{ <i>Pirenella cingulata</i> , Spionidae sp.}	{ <i>Clithon retropictum</i> }	0.005	1.000	0.005 </td <td>4.467</td> <td>1</td>	4.467	1

LHS (Left-Hand Side): This column shows combinations of conditions or entities, RHS (Right-Hand Side): This is the outcome or effect of the conditions in the LHS, Support: Indicates the frequency of the rule in the dataset, Confidence: Shows the probability of the RHS occurring when the LHS is true, Coverage: The proportion of the dataset that contains the LHS, Lift: A measure of the strength of the association between the LHS and RHS. A lift greater than 1 indicates a strong association, Count: Number of times the rule is observed in the dataset.

류(*Melitidae* sp.), 깔따구류(*Chironomidae* sp.)의 조합이 기수갈고등과 관련된 것으로 나타나, 약 8.46%의 지지도와 51.52%의 신뢰도를 기록했다. 이는 기수 및 해수역에 서식하는 종들과 달리 담수에 출현하는 실지렁이, 깔따구류를 포함한 조합이 높은 지지도와 신뢰도를 보이며, 해수와 담수가 만나는 기수역의 특성을 나타냈다. 다른 주요 규칙들은 *Nereidae* sp., *L. gotoi*, *Melitidae* sp., *H. penicillatus*, *Chironomidae* sp.의 조합이 기수갈고등과 연관되어 약 7.96%의 지지도를 보였으며, *A. japonica*와 *H. penicillatus*의 조합은 6.97%의 지지도를 나타냈다.

연관 규칙 분석에서 향상도 기준으로 상위 50개의 규칙을 살펴본 결과, 기수갈고등과 연관된 다양한 생물종의 조합이 나타났다(Table 2, Fig. 4b). 이들 규칙은 모두 향상도가 4.47로 매우 높으며, 모든 규칙에 대한 신뢰도는 100%였다. 이러한 높은 향상도는 특정종의 조합이 기수갈고등의 출현과 매우 밀접하게 관련되어 있음을 나타낸다. 예를 들어, 둥근입기수우렁이(*Stenothyra glabrata*)와 비틀이고등(*Pirenella cingulata*)의 조합이 기수갈고등과 연관된 것으로 나타났으며, 이 규칙은 전체 관찰 중 0.5%의 지지도를 보였다. 비슷한 조합으로는 비틀이고등(*P. cingulata*)과 백금갯지렁이류(*Nephthyidae* sp.), 동다리(*Cerithidea rhizophorarum*)와 비틀이고등(*P. Cingulata*) 등이 있으며, 이들 역시 동일한 향상도와 신뢰도를 나타냈다.

연관 규칙 분석의 결과를 카테고리별로 분석한 데이터는 ‘High Lift and High Confidence’와 ‘High Lift but not High Confidence’의 두 가지 카테고리로 구분되었다(Fig. 5). 분석된 총 2,306,452개의 규칙 중에서 ‘High Lift and High Confidence’ 카테고리에 속하는 규칙의 수는 2,263,683개로, 전체 규칙의 약 98.19%를 차지했다. 이는 이 카테고리의 규칙들이 전체 데이터 세트에서 매우 높은 향상도와 신뢰도를 가지고 있음을 나타낸다. 반면, ‘High Lift but not High Confidence’ 카테고리에 속하는 규칙은 41,769개로 전체의

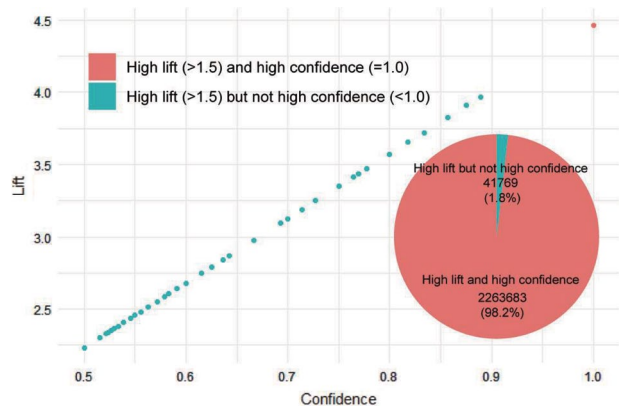


Fig. 5. Comprehensive analysis of rule characteristics and category proportions: The main scatter plot displays association rules classified by lift and confidence. The inset pie chart shows the proportional distribution of rule categories.

약 1.81%를 차지했다.

4. K-means 군집분석 및 계층적 군집분석

K-means 군집분석 결과 총 5개의 군집으로 구분되었으며, 이 중 가장 큰 군집 5는 72종 중 64종을 포함하고 있어, 전체 종의 약 88.89%를 차지한다(Fig. 6a). 이는 대부분의 종들이 유사한 생태적 특성을 공유하고 있음을 나타낸다. 기수갈고등 역시 군집 5에 포함되어 있다. 나머지 군집들은 상대적으로 적은 종을 포함하고 있으며, 그 크기는 각각 군집 1에 1종(1.39%), 군집 2에 4종(5.56%), 군집 3에 1종(1.39%) 그리고 군집 4에 2종(2.78%)으로 구성되었다. 계층적 군집분석을 통해 형성된 4개의 군집 중, 군집 1은 가장 많은 종을 포함하고 있으며, 기수갈고등을 포함한 대다수 종(63종, 87.50%)이 이 군집에 속한다. 나머지 군집들은 소수의 종으로 구성되어 있으며, 각각 1종 또는 2종으로 이루어져 있다(Fig. 6b).

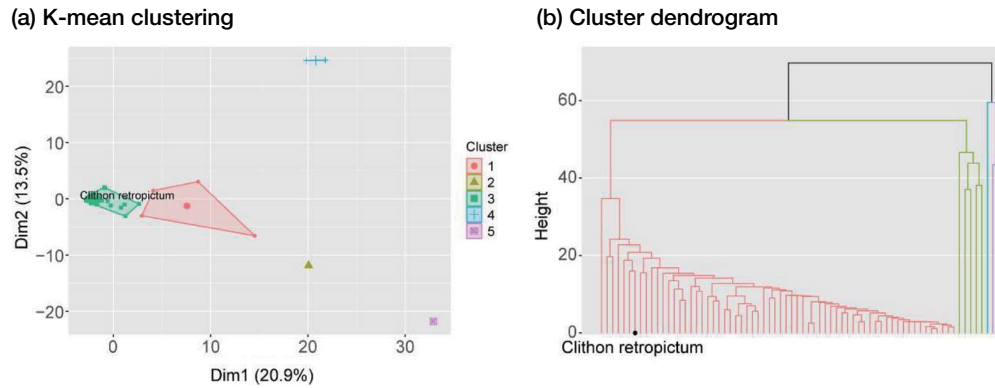


Fig. 6. Comparative analysis of species distributions: K-means (a) and Hierarchical clustering (b).

고찰

1. 연관 규칙 학습을 통한 기수갈고둥을 중심으로 한 생물종 간 공동출현

이 연구는 연관 규칙 학습과 군집분석을 활용하여 기수갈고둥과 다른 종 사이의 공동출현을 확인했다. 연관 규칙 분석 결과, 특히 기수우렁이와 같은 종과 기수갈고둥 사이의 공동출현이 높은 지지도와 신뢰도를 나타내며, 기수갈고둥의 생태계 내 중요성을 강조한다. 기수우렁이의 출현이 기수갈고둥의 생존 및 번식에 긍정적인 조건을 제공하는 것으로 나타나, 이러한 공동출현은 기수갈고둥이 특정 환경 조건, 특히 기수역과 같은 생태계에서 서식할 수 있음을 나타낸다. 기수우렁이와의 공동출현을 중심으로, 기수갈고둥이 선호하는 환경 조건을 더욱 명확히 파악할 수 있다. 이 종은 주로 서해와 남해 해안가의 기수역에서 발견되며, 이러한 지역은 기수갈고둥이 번식하고 생존하기에 적합한 서식지를 제공한다. 또한, 연구 결과는 기수갈고둥과 공동출현하는 다른 종들이 기수갈고둥의 서식지 선택에 영향을 줄 수 있음을 나타낸다. 이는 기수갈고둥의 서식 가능성을 평가하는 데 있어서 생물학적 공동출현을 고려하는 것이 중요함을 의미한다. 멸종위기 야생생물처럼 관찰 빈도가 낮은 종을 직접적으로 조사하지 않고도 연관 규칙 학습에서 확인된 공동출현이 높은 종 중 쉽게 관찰 가능한 종을 연구함으로써 멸종위기야생생물의 서식환경을 이해하는 데 도움이 될 수 있다.

많은 규칙들이 높은 향상도와 신뢰도를 보였음에도 불구하고, 지지도가 상대적으로 낮은 것도 관찰되었다. 이는 기수갈고둥과 연관된 종들이 기수갈고둥이 출현한 서식환경이나 조건에서는 강한 연관성을 가지지만, 전체 생태계 내에서는 그들의 출현 빈도가 낮음을 의미한다. 이는 기수갈고둥이 특정 조건에서만 번성할 수 있으며, 이러한 조건의 변화

가 그들의 생존에 직접적인 영향을 미칠 수 있음을 나타낸다 (Grüneberg and Nugaliyadde, 1976).

종이 다양한 다른 종과 연관되어 있다는 것은 그 종이 특정한 생태적 틈새나 조건에 잘 적응하고 있음을 나타낼 수 있다(Han *et al.*, 2021; Jang *et al.*, 2021). 하지만 다양한 종과의 관계는 그 종이 어떤 유형의 서식지에서 발견될 수 있는지에 대한 정보를 제공하지만, 그 종이 넓은 지리적 범위에 걸쳐 존재한다는 것을 의미하지 않는다. 기수갈고둥은 하구 생태계에서 중요한 역할을 할 수 있지만, 이러한 역할은 특정 생태계나 조건에 국한될 수 있다(Shigemiy and Kato, 2001). 서식지의 특성화와 관련하여, 연관 규칙 학습 결과는 서식지의 특성을 더 잘 이해하는 데 도움이 될 수 있다.

2. 연관 규칙 학습과 전통적인 군집분석 방법론의 비교

연관 규칙 학습과 전통적인 군집분석 방법론은 생물종들 간의 연관성을 이해하는 데 서로 다른 접근 방식을 제공하는데, 이 차이점은 생물종 간의 관계를 파악하고 생태계 연구에 새로운 정보를 제공하는 데 중요하다(Legendre and Fortin, 1989). 연관 규칙 학습은 대규모 데이터 세트 내에서 종들 간의 동시 출현 패턴을 분석하여 생물종들 간의 연관성을 파악한다. 이 방법은 종들 간의 관계를 직접적으로 밝히며, 특정 환경 조건이나 다른 생물종들의 관계에서 종의 출현 패턴을 분석하는 데 유용하다. 각 규칙의 지지도, 신뢰도, 향상도 등을 통해 종들 간 연관성의 강도와 중요성을 정량적으로 평가할 수 있어, 생태계 내 복잡한 종 간 관계를 이해하는 데 효과적이다. 반면 전통적인 군집분석 방법론은 데이터 세트 내에서 유사한 특성을 가진 종들을 그룹화하여, 생태계 내에서 서식지 또는 행동 유사성에 기반한 군집을 형성한다(Green, 1980; Wootton, 1994).

K-means 군집분석과 계층적 군집분석은 이러한 군집을

식별하는 데 사용되며, 종들 간의 유사성과 차이를 기반으로 생태적 관계를 이해하는 데 도움을 준다. 이 방법은 생태계 내에서 종들이 어떻게 그룹화되는지를 파악하는 데 중요하지만, 종들 간의 직접적인 연관성을 파악하기에는 한계가 있다. 본 연구에서는 이 두 방법론을 병행하여 사용함으로써, 기수갈고등을 포함한 공동으로 출현하는 다양한 종들 간의 관계를 더 깊이 이해할 수 있었다. 연관 규칙 학습은 종들 간의 직접적인 연관성을 밝히는 반면, 군집분석은 종들이 서식하는 환경 및 행동 패턴에 따른 유사성을 분석했다. 이러한 다각적인 접근은 생태계 내 복잡한 종 간의 관계를 보다 포괄적으로 이해하는 데 기여할 것으로 생각된다(Brown *et al.*, 2001; Carrara *et al.*, 2015).

그럼에도 불구하고 연관 규칙 학습을 통해 생물과 생물 또는 생물과 환경 간의 관계를 추론하는 데 문제점도 있었다. 연관 규칙 학습은 대규모 데이터 세트에서 효과적인 방법이지만, 데이터의 복잡성과 다양성에 따라 해석이 어려울 수 있다(Ünvan, 2021). 특히 다양한 환경 조건이나 상호작용을 포함하는 복잡한 생태계 데이터에서는 단순한 규칙이 실제 생태계의 동적 관계를 완전히 반영하지 못할 위험이 있다. 이와 대조적으로 군집분석은 종의 유사성과 차이를 기반으로 분류하므로, 생태적 특성이나 행동 패턴을 반영하는 데 제한적일 수 있다. 생태계 내에서 종들의 공동출현은 시간적 및 공간적으로 변동될 수 있으며, 연관 규칙 학습은 주로 정적인 데이터 패턴 분석에 초점을 맞추는 경향이 있다. 생태계는 지속적으로 변화하고 적응하기 때문에, 군집분석 역시 이러한 동적 변화를 반영하는 데 한계가 있을 수 있다. 따라서 두 방법을 병행할 때 시간적 및 공간적 요소를 고려하는 것이 중요하다. 방법론의 통합과 발전을 통해 두 방법론의 한계를 극복하고 보다 정교한 생태계 분석이 가능하다. 연관 규칙 학습으로 도출된 연관성 패턴을 군집분석과 결합하면, 생태계 내에서 종들이 어떻게 공동으로 출현하며 이 출현이 어떻게 시간에 따라 변화하는지 보다 깊이 이해할 수 있다. 이러한 통합 접근법은 생태계 연구에 있어 보다 포괄적이고 동적인 관점을 제공하며, 생태계의 복잡성을 더 잘 이해하는 데 기여할 수 있다(Smith *et al.*, 2017; Leote *et al.*, 2020).

적 요

본 연구는 한국의 멸종위기야생생물 II급으로 분류된 기수갈고등(*Clithon retropictum*)과 생태계 내 다른 종들 사이의 공동출현 양상을 분석하고자, 연관 규칙 학습과 군집분석 방법론을 통합적으로 사용했다. 이 연구의 주요 목적은 기수갈고등과 다른 생물종 사이의 공동출현 패턴을 파악하는 것이

다. 대규모 데이터 세트를 통한 연관 규칙 학습으로 종 간의 공동 출현 패턴과 그들사이의 연관성을 탐색하였고, K-mean 와 계층적 군집분석을 통해 종들 간의 생태적 유사성과 차이에 기반으로 한 그룹화를 진행하였다. 연구 결과는 기수갈고등과 다른 종들 사이에 상당한 공동 출현 관계가 있음을 보여주며, 이러한 공동 출현은 특정 환경 조건과 밀접하게 연결되어 있음을 나타냈다. 특히, 기수갈고등이 출현하는 지점에서는 특정종의 출현 빈도가 높게 나타나는 패턴을 보였고, 이는 기수갈고등이 생태계 내에서 중요한 생물지표종으로서의 역할을 수행하고 있음을 시사한다. 이러한 접근 방식은 기수갈고등과 같은 멸종 위기종의 보존과 생태계 관리 전략 수립에 있어 중요한 의미를 지닌다. 생태계 내에서 공동으로 출현하는 종들 간의 복잡한 연관성을 이해함으로써, 보다 효과적인 보존 전략을 개발하고 생태계의 건강과 안정성을 유지하는 데 기여할 수 있다. 본 연구는 생태계 연구에 있어 데이터 기반 접근법의 중요성을 강조하며, 생물 다양성 보존을 위한 새로운 방향을 제시한다.

저자정보 임성호((주)생태조사단 부장), 도윤호(공주대학교 교수)

저자기여도 개념설정: 임성호, 도윤호, 자료수집: 임성호, 분석: 임성호, 도윤호, 원고작성: 임성호, 도윤호

이해관계 이 논문에는 이해관계 충돌의 여지가 없습니다.

연구비 본 결과물은 환경부의 재원으로 국립환경과학원의 지원을 받아 수행하였습니다(NIER-2023-04-02-097).

REFERENCES

- Aronne, G., M. Giovanetti, M.R. Guarracino and V. De Micco. 2012. Foraging rules of flower selection applied by colonies of *A. pis mellifera*: ranking and associations of floral sources. *Functional Ecology* **26**: 1186-1196.
- Brown, J.H., T.G. Whitham, S.K. Morgan Ernest and C.A. Gehring. 2001. Complex species interactions and the dynamics of ecological systems: long-term experiments. *Science* **293**: 643-650.
- Carrara, F., A. Giometto, M. Seymour, A. Rinaldo and F. Altermatt. 2015. Inferring species interactions in ecological communities: a comparison of methods at different levels of complexity. *Methods in Ecology and Evolution* **6**: 895-906.
- Cognetti, G. and F. Maltagliati. 2000. Biodiversity and adaptive mechanisms in brackish water fauna. *Marine Pollution Bulletin* **40**: 7-14.
- Dale, V.H. and S.C. Beyeler. 2001. Challenges in the development and use of ecological indicators. *Ecological Indicators* **1**:

- 3-10.
- Green, R.H. 1980. Multivariate approaches in ecology: the assessment of ecologic similarity. *Annual Review of Ecology and Systematics* **11**: 1-14.
- Grüneberg, H. and L. Nugaliyadde. 1976. Population studies on a polymorphic prosobranch snail (*Clithon* (*Pictoneritina*) *oualaniensis* Lesson). *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **275**: 387-429.
- Hahsler, M. 2017. ArulesViz: interactive visualization of association rules with R. *The R Journal* **9**: 163.
- Hahsler, M., B. Grün and K. Hornik. 2005. arules - A computational environment for mining association rules and frequent item sets. *Journal of Statistical Software* **14**: 1-25.
- Han, S.P., I.C. Hwang and S.J. Kwon. 2021. Studies on Distribution and Ecology of *Clithon retropictus* (Martens, 1879) in South Korea. *Journal of Wetlands Research* **23**: 317-326.
- Jang, K.G., M.S. Kim, S.G. Jo, W. Yih and H.S. Kim. 2021. Habitat Environments and Spatiotemporal Distribution of *Clithon retropictum* at the Estuaries of Bonggang and Miryoung Streams in the Southern Coast of Korea. *Ocean and Polar Research* **43**: 127-140.
- Keil, P., D. Storch and W. Jetz. 2015. On the decline of biodiversity due to area loss. *Nature Communications* **3**: 8837.
- Legendre, P. and M.J. Fortin. 1989. Spatial pattern and ecological analysis. *Vegetatio* **80**: 107-138.
- Leote, P., R.L. Cajaiba, J.A. Cabral, A.D. Brescovit and M. Santos. 2020. Are data-mining techniques useful for selecting ecological indicators in biodiverse regions? Bridges between market basket analysis and indicator value analysis from a case study in the neotropics. *Ecological Indicators* **109**: 105833.
- Lindenmayer, D.B. and G.E. Likens. 2010. The science and application of ecological monitoring. *Biological Conservation* **143**: 1317-1328.
- Maechler, M. 2018. Cluster: cluster analysis basics and extensions. R package version 2.0. 7-1.
- MOE. 2016. Water quality monitoring program. The Ministry of Environment, Korea.
- Park, W.B., S.H. Lim, D.H. Won, K.L. Lee, C. Hong and Y. Do. 2022. Occupancy Probability Estimation of Endangered Species *Clithon retropictus*. *Korean Journal of Ecology and Environment* **55**: 76-83.
- Pedersen Zari, M., M. Mackinnon, K. Varshney and N. Bakshi. 2022. Regenerative living cities and the urban climate-bio-diversity-wellbeing nexus. *Nature Climate Change* **12**: 601-604.
- R Core Team, R. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Samecka-Cymerman, A., A. Stankiewicz, K. Kolon, A.J. Kempers and R.S.E.W. Leuven. 2010. Market basket analysis: a new tool in ecology to describe chemical relations in the environment - A case study of the fern *Athyrium distentifolium* in the Tatra National Park in Poland. *Journal of Chemical Ecology* **36**: 1029-1034.
- Shaw, R.C., A.L. Greggor and J.M. Plotnik. 2021. The challenges of replicating research on endangered species. *Animal Behavior and Cognition* **8**: 240-246.
- Shigemiyu, Y. and M. Kato. 2001. Age distribution, growth, and lifetime copulation frequency of a freshwater snail, *Clithon retropictus* (Neritidae). *Population Ecology* **43**: 133-140.
- Shogren, J.F., J. Tschirhart, T. Anderson, A.W. Ando, S.R. Beissinger, D. Brookshire, G.M. Brown Jr, D. Coursey, R. Innes and S.M. Meyer. 1999. Why economics matters for endangered species protection. *Conservation Biology* **13**: 1257-1261.
- Smith, D.I., M.F. Curran and A.V. Latchininsky. 2017. Market basket analysis of grasshopper (Orthoptera: ACRIDIDAE) assemblages in eastern Wyoming: a 17-year case study using associative analysis for ecological insights into grasshopper outbreaks. *Ecological Entomology* **42**: 379-382.
- Souza, C.R., V.A. Maia, N. Aguiar-Campos, C.L. Farrapo and R.M. Santos. 2021. Tree species consistent co-occurrence in seasonal tropical forests: an approach through association rules analysis. *Forest Systems* **30**: e006.
- Szaro, R.C. 2008. Endangered species and nature conservation: science issues and challenges. *Integrative Zoology* **3**: 75-82.
- Thorp, J.H., M.C. Thoms and M.D. Delong. 2006. The riverine ecosystem synthesis: biocomplexity in river networks across space and time. *River Research and Applications* **22**: 123-147.
- Ünvan, Y.A. 2021. Market basket analysis with association rules. *Communications in Statistics-Theory and Methods* **50**: 1615-1628.
- Wootton, J.T. 1994. The nature and consequences of indirect effects in ecological communities. *Annual Review of Ecology and Systematics* **25**: 443-466.