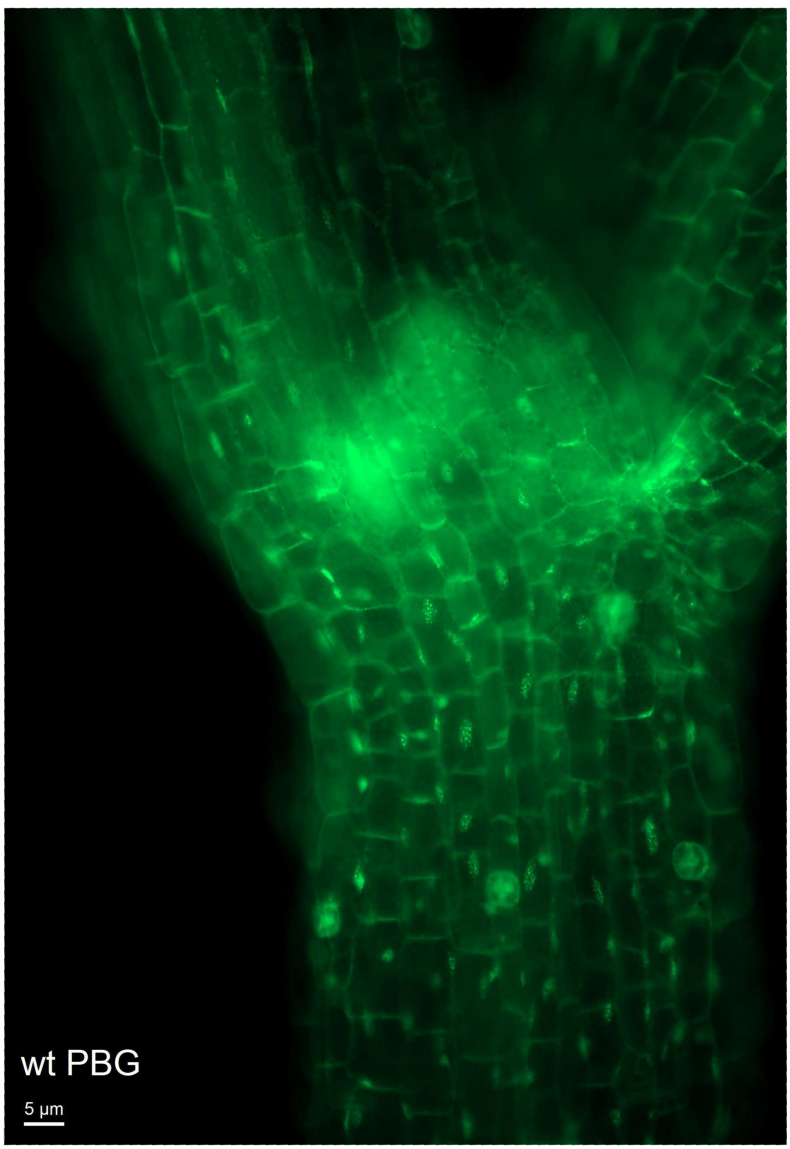
STUDIA UNIVERSITATIS
BABEŞ-BOLYAI

UNIVERSITATEA BABEŞ-BOLYAI
BABEŞ-BOLYAI TUDOMÁNYEGYETEM
BABEŞ-BOLYAI UNIVERSITÄT
BABEŞ-BOLYAI UNIVERSITY
TRADITIO ET EXCELLENTIA



BIOLOGIA

Volume 71 (2026), June, No. 1

STUDIA
UNIVERSITATIS BABEŞ-BOLYAI
BIOLOGIA

1 / 2026

January – June

ISSN (online): 2065-9512 | ISSN-L: 1221-8103

Published by Babeş-Bolyai University.

EDITORIAL BOARD
STUDIA UNIVERSITATIS BABEŞ-BOLYAI BIOLOGIA

YEAR **Volume 71 (LXXI), 2026**
MONTH **JUNE**
ISSUE **1**

PUBLISHED ONLINE: 2026-06-25
ISSUE DOI: 10.24193/subbbiol.2026.1
ISSN (online): 2065-9512
https://studiabiologia.reviste.ubbcluj.ro/

STUDIA UNIVERSITATIS BABEŞ-BOLYAI

BIOLOGIA

1

SUMAR – CONTENT – SOMMAIRE – INHALT

RESEARCH

All authors are responsible for submitting manuscripts in comprehensible US or UK English and ensuring scientific accuracy.

Original picture on front cover: ©Leonard Calistru, Florent Velay, Dorina Podar, Robert Blanvillain. Confocal image of an *Arabidopsis thaliana* wild-type seedling expressing PHYB–GFP (wt PBG), showing the distribution of the fluorescent fusion protein and its accumulation in nuclear photobodies.

STUDIA UNIVERSITATIS BABEŞ-BOLYAI BIOLOGIA, 71, 1, 2026 (p. 5-16)
doi:10.24193/subbbiol.2026.1.01

First record of *Parasyrphus malinellus* (Collin, 1952) (Diptera: Syrphidae) in Romania at low altitude in the Danube Gorge

Amalia-Raluca Dumbravă[1,2]✉, Andreea-Maria Lazăr[3], Alina-Florentina Muşet[3] and Tamara Tot[4]

[1] University of Oradea, Doctoral School of Biomedical Sciences, Oradea, Romania; [2] Iron Gates Natural Park, Orşova, Romania; [3] University of Oradea, Faculty of Informatics and Sciences, Department of Biology, Oradea, Romania; [4] University of Novi Sad, Faculty of Sciences, Department of Biology and Ecology, Novi Sad, Serbia;
✉***Corresponding author, E-mail: amalia.dumbrava@pnportiledefier.ro***

Article history: Received 3 February 2026; Revised 26 March 2026; Accepted 4 April 2027; Available online 25 June 2026

Abstract. Hoverflies (Diptera: Syrphidae) represent an ecologically important group of pollinating insects. Of these, *Parasyrphus malinellus* (Collin, 1952) is a Palaearctic species, mainly distributed in northern Europe. In southern Europe, particularly on the Balkan Peninsula, its distribution is scattered and has been recorded only recently at high altitudes. Although Romania lies north of the Balkans, *P. malinellus* has not previously been recorded in the country. In April 2025, we recorded a female in the Mraconia Valley Basin (Danube Gorge, southwestern Romania). This is the first record of *P. malinellus* for the Romanian Carpathians and the country. This record partially fills the distribution gap between the species' known distribution in northern and western Europe and its southern Balkan occurrences (Serbia and Greece). Unlike other records from the Balkan Peninsula, where *P. malinellus* is present at high elevations, we recorded it at an unexpectedly low altitude, at only approximately 400 m a.s.l. Moreover, although *P. malinellus* is typically associated with coniferous forests, it was found in a humid valley surrounded by steep slopes dominated by beech and oak forests, in a region where conifers are not naturally present.

Thus, *P. malinellus* is another mountain species, probably relict, present in the Danube Gorge close to its southern distribution limit, at a low altitude and in atypical habitats. This finding further supports the role of the Danube Gorge as a refugium, increasing its conservation value. At the same time, it highlights the limited knowledge of hoverfly diversity in the Carpathian Mountains, even within well-studied protected areas.

Keywords: hoverflies, low altitude, forest, habitats, refugium, distribution record.

Introduction

Pollinating insects are crucial to both natural and agricultural ecosystems (see Takov *et al.,* 2025). An important group of pollinators that also provides numerous other ecosystem services is hoverflies (Diptera, Syrphidae) (Doyle *et al.,* 2020). However, the decline of these pollinating insects appears to have accelerated in recent years in some regions of western Europe, such as the Netherlands (Zeegers *et al.,* 2024). This fact is also true in other areas, as in the Carpathian region from central Europe, hoverflies are the insect group that has undergone the most significant abundance decline due to climate change (Haris *et al.,* 2025). Even in areas from south-eastern Europe where hoverflies had a relatively favorable status, these insects face many threats (Milosavljević *et al.,* 2026). In the Danube Gorge (south-western Romania), 111 syrphid species belonging to 44 genera were recorded (Dumbravă & Cupșa, 2025), representing almost a fourth of the total number of species from the country, namely 419 (Reverté *et al.,* 2023). However, the inventory remains incomplete, primarily due to the region's large area, which represents a Natural Park (Dumbravă & Cupșa, 2025). At the same time, Romania is still poorly represented in terms of the abundance of hoverfly records compared with other regions, especially in western Europe (Sentil *et al.,* 2026).

From the genus *Parasyrphus* Matsumura, 1917 eight species have been recorded in Romania (Stănescu & Pârvu, 2005), but in the Romanian Danube Gorge only one species was previously mentioned (Dumbravă & Cupșa, 2025). From this genus, *P. malinellus* (Collin, 1952) is a Palaearctic hoverfly species characteristic of northern and central Europe, with a distribution ranging from western Europe to Siberia, while reaching the Alps at the southern edge of its range (Reemer, 2009; Speight, 2024). The species is well represented in northern and western Europe, including Belgium, the Netherlands, Norway, Finland, and Sweden (Verlinden, 1994; Reemer, 2009; Hågvar & Nielsen, 2007; Nielsen &

Svendsen, 2014; Haarto & Kerppola, 2014; van Steenis, 2011). To the south, *P. malinellus* reaches into Slovenia (de Groot & Govedič, 2008; van Steenis *et al.,* 2013) or into southern France (Lebard & Speight, 2019) in the Mediterranean area, but in the Alps region, where its presence is known (Speight, 2024). Recently, it was identified in the southern part of the Balkan Peninsula, in northern Greece, near the border with Bulgaria, where it is very localized, as it is present only in an area with spruce forests, at 1550 m altitude, in the Rhodope Mountains (Van de Meutter *et al.,* 2025). In Serbia, it is found in montane coniferous habitats above 1300 m (Vujić et al., 2018; *Parasyrphus malinellus* (Collin, 1952) in GBIF Secretariat, 2023), and in Hungary it occurs in higher northern and western forested regions (Tóth, 1995, 2011). It occupies coniferous forests (Löhr, 1990; Reemer, 2009; Speight, 2004; 2024), but it also appears sometimes in deciduous forests (Nilsson *et al.,* 2007; Reemer, 2009). In Romania, *P. malinellus* has not been reported previously, not only in the Danube Gorge (Dumbravă & Cupșa, 2025) but anywhere in the country (Brădescu, 1991; Stănescu & Pârvu, 2005). Thus, herein we report the first record of *P. malinellus* in Romania, from the Danube Gorge (Mraconia River basin), and describe its occurrence at an unusually low altitude and in an unusual habitat. However, the Danube Gorge is known to shelter mountain species at low altitudes (e.g., Paşcovschi, 1956; Covaciu-Marcov *et al.,* 2009, 2025; Teodor *et al.,* 2019; 2025; Petruș-Vancea *et al.,* 2024). The region's particularities are also clear for hoverflies (Dumbravă & Cupșa, 2025). Thus, our study also highlights the ecological and biogeographical importance of the presence of *P. malinellus* in the region and underlines the need for further research on hoverflies in the Romanian Carpathians.

Materials and methods

Fieldwork was conducted at the end of April 2025 in the Romanian Danube Gorge region, while also investigating other insect groups in the area (Teodor *et al.,* 2025). Investigations included the Mraconia River basin in the eastern sector of the Danube Gorge (Iron Gates Natural Park), situated in the Almăj Mountains. Along the Muşchiosul Mare tributary, a 2 km transect was surveyed along a forest road, and hoverflies were collected directly from flowering plants using hand nets, as previously (Dumbravă & Cupșa, 2025).

Collected specimens were preserved in ethanol and identified to species level in the laboratory under a stereomicroscope using the following references: van Veen (2004), Bartsch *et al.* (2009), and Speight & Sarthou (2017). This species closely resembles *P. proximus* (e.g., Speight, 2024) and is difficult to distinguish from other congeners (Löhr, 1990; van Steenis, 2011). Identification was confirmed using diagnostic characters indicated in the literature (van Veen, 2004; Bartsch

et al., 2009; Speight & Sarthou, 2017). Tergites 3 and 4 are black with distinct transverse yellow bands. The basoflagellomere is black, with yellow ventrally. Sternites 3 and 4 are yellow, each with a large triangular black marking. In males, about one-eighth of the hind femur is yellow, while the fore and mid tarsi are dark greyish. Females have mostly black hind femora. This species is morphologically very similar to *P. proximus* Mutin, 1991, from which it can be distinguished by the shape and pattern of dark markings on the sternites and the coloration of the lower part of the male face. In *P. malinellus*, sternites 3 and 4 have triangular black markings on the posterior half, whereas in *P. proximus*, these segments bear black bands. Male *P. malinellus* have a broadly black mouth edge, while male *P. proximus* have a narrowly black mouth edge, sometimes yellowish anteriorly (characters after van Veen, 2004).

Results

On 25 April 2025, a single *P. malinellus* specimen was collected from the Danube Gorge region, in the Mraconia River basin (Figure 1). The specimen was an adult female, identifiable by the pointed abdomen and dichoptic eyes (Figure 2). It was found at approximately 400 m altitude, near the Romania–Serbia border, about 8 km in a straight line from the Danube. The indicated altitude is approximate, because while walking the 2-kilometer-long transect on Mușchiosul Mare tributary, we collected numerous hoverflies, which were only subsequently determined in the laboratory. Thus, it is impossible to know which individual was captured at what meter of the transect. The hoverflies were personally collected and determined by the authors. The collected specimen is stored in the first author's personal collection. The valley is narrow and humid, with steep slopes dominated by beech (*Fagus* spp.) and oak (*Quercus* spp.) forests. *P. malinellus* was observed in open and sunny microhabitats along the forest road, in rocky areas adjacent to the water, and on concrete walls stabilizing the road edges. These locations contained numerous flowering plants (*Ranunculus repens, Tussilago farfara, Anemone nemorosa, Lamium galeobdolon, Lamium maculatum, Urtica dioica*, etc.), which likely attracted the specimen.

Discussion

In the Danube Gorge region, *P. malinellus* was recorded for the first time in Romania, as it had not been previously mentioned in the country (Brădescu, 1991; Stănescu & Pârvu, 2005). The Romanian Carpathians likely provide suitable habitats for this species, which is associated with coniferous forests (Speight,

2024; Reemer, 2009; Löhr, 1990; Hågvar & Nielsen, 2007). The species was not observed in the high-altitude, humid spruce forests of the Carpathians but was found in the Danube Gorge, one of the warmest regions of southwestern Romania (Gheorghe *et al.,* 2025), at approximately 400 m altitude. This is considerably lower than altitudes reported elsewhere in the Balkan Peninsula (*Parasyrphus malinellus* (Collin, 1952) in GBIF Secretariat, 2023, Van de Meutter *et al.,* 2025), suggesting that the species can persist at low elevations under favorable microclimatic conditions. It is well known that the Danube Gorge shelters certain mountain species at low-altitudes, both among plants (Csürös *et al.,* 1968; Schneider-Binder, 2014; 2016; Petruș-Vancea *et al.,* 2024) and animals (Iftime, 2005; Cicort-Lucaciu *et al.,* 2017; Covaciu-Marcov *et al.,* 2009, 2025; Teodor *et al.,* 2019). Even nowadays, such species, considered potentially glacial relicts, are identified in the region (Teodor *et al.,* 2025). In the Mraconia valley, plant species and associations were recorded at their lowest altitude in the Carpathians (Schneider-Binder, 2016). Also, in the case of hoverflies, mountain species were recorded at low altitudes alongside southern, Mediterranean species (Dumbravă & Cupșa, 2025). As in other cases (Teodor *et al.,* 2025), the record of *P. malinellus* in the Danube Gorge seems to partially fill the gap between populations in the Balkan Peninsula and those situated north and west of Romania (Tóth, 2011; Żoralski & Kowalczyk, 2015; Mazánek *et al.,* 2025). The specimen was collected at the end of April, consistent with the species' known flight period (April–July; Speight, 2024).

In the upper sector of Mraconia valley, the annual average temperature is lower than in most of the Danube Gorge region (Stoenescu *et al.,* 1966; Gheorghe *et al.,* 2025), thus, probably this valley functions as a microclimate refugia for cold climate species. This is important in the context of the intense heat waves of recent years, as evidenced in certain mountain regions in Europe, where those microrefugia could disconnect from the macroclimatic trends of the surrounding regions (Finocchiaro *et al.,* 2024). A similar situation was previously reported in the neighboring Eșelnița valley, whose colder microclimate is likely responsible for the presence of *Vaccinium myrtillus* at very low altitudes (Petruș-Vancea *et al.,* 2024). This phenomenon is likely widespread in the Danube Gorge region and is responsible for the co-occurrence of warm-climate and mountainous species at low altitudes (e.g., Pașcovschi, 1956; Covaciu-Marcov *et al.,* 2009). Moreover, in recent years in the region, both mountain species associated with a colder climate (Petruș-Vancea *et al.,* 2024; Covaciu-Marcov *et al.,* 2025; Teodor *et al.,* 2025) and Mediterranean species (Ruicănescu & Dumbravă, 2020; Dumbravă & Ruicănescu, 2023; Manci & Dumbravă, 2026) have been recorded. Probably the narrow valleys in the Danube Gorge preserved from the last glacial maximum glacial and postglacial relicts, as other remarkable habitats from central Europe (Słowińska *et al.,* 2026), a fact mentioned as possible also in the case of other lowland insect populations in Romania (Sitar *et al.,* 2025). Even in the

case of hoverflies, climatic peculiarities could generate a local microclimate that could ensure the survival of residual populations of some northern species (Burgio *et al.,* 2025).

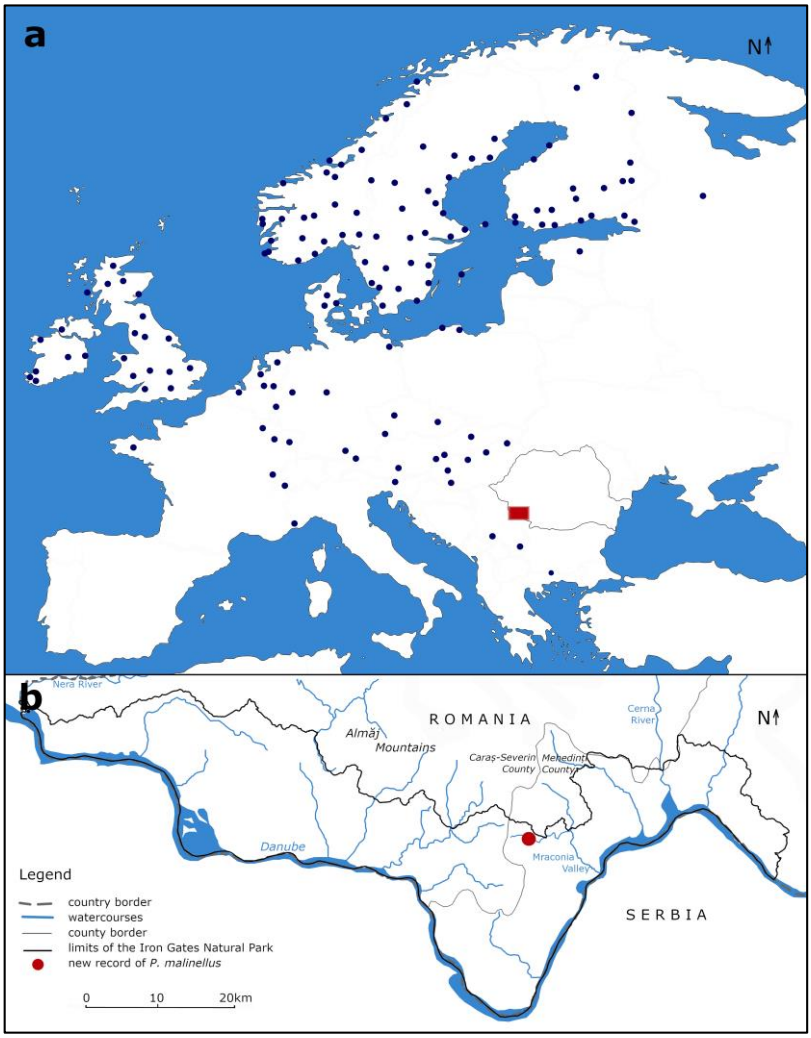


Figure 1. a. Distribution of *P. malinellus* in Europe, according to data available on GBIF (*Parasyrphus malinellus* (Collin, 1952) in GBIF Secretariat, 2023) and other sources (Tóth, 2011; Żoralski & Kowalczyk, 2015; Mazánek *et al.,* 2025; Van de Meutter *et al.,* 2025; in some cases a point represents an entire region, not a specific occurrence record); **b.** first record of *P. malinellus* in Romania

The habitat where *P. malinellus* was found is also notable. Although primarily associated with coniferous forests (Löhr, 1990; Reemer, 2009; Speight, 2004, 2024), it can occasionally occur in deciduous forests (Nilsson *et al.,* 2007; Reemer, 2009). At the southern edge of its range in Greece, it occurs only in coniferous forests (Van de Meutter *et al.,* 2025). In contrast, in the Danube Gorge, there are no coniferous forests; only beech and oak forests (e.g., Reif *et al.,* 2022), except for the black pine (Schneider-Binder, 2014), which is anyway very rare in the gorge, and to our best knowledge absent from the Mraconia valley. The only coniferous trees from the Mraconia basin are a few small-sized and relatively recent plantations. This suggests that *P. malinellus* may be a relict of ancient coniferous forests that once occurred in the Banat Mountains, surviving the extinction of native conifers after the last glacial maximum (Paşcovschi, 1967; Rösch & Fischer, 2000) and subsequently adapting to deciduous habitats and present coniferous plantations. This seems plausible, as the southwestern Carpathians were considered a potential glacial refugium for certain species (e.g., Paşcovschi, 1967; Fijarczyk *et al.,* 2011; Magyari *et al.,* 2012; Şuteu *et al.,* 2023; Voisin *et al.,* 2026).

Figure 2. *Parasyrphus malinellus*: **a, d.** female, habitus ventral view; **b.** head lateral view; **c.** habitus dorso-lateral view.

The microhabitat characteristics, sunny edges inside forests, align with the species' ecological requirements (Reemer, 2009). Although widely distributed in northern Europe, *P. malinellus* remains rare throughout its range (Bartsch *et al.,* 2009). Thus, probably it was not previously mentioned in Romania because of its general rarity and largely arboreal life (Speight, 2024), even if the hoverflies on the Danube Gorge were recently studied (Dumbravă & Cupșa, 2025), while the neighboring areas in the Southern Carpathians were studied in the past (Brădescu, 1975-1976; 1986; 1995). Nevertheless, studies on hoverflies in Romania are likely insufficient, as the entire Balkan region is understudied (Reverté *et al.,* 2023). The recent first-time identification of another insect species in Romania was also attributed to the low number of entomological studies (Iacob *et al.,* 2021). Also, the population could be isolated and thus difficult to identify, as the large surface area of the protected areas has repeatedly been considered an impediment to faunistic studies (Covaciu-Marcov *et al.,* 2025; Dumbravă & Cupșa, 2025). This discovery highlights the need for further research in the Romanian Carpathians to determine whether the Danube Gorge population is isolated or whether other populations remain unrecorded due to insufficient study.

Acknowledgements. Our study was made in collaboration and with the support of the Iron Gates Natural Park Administration.

References

Bartsch, H., Binkiewicz, E., Råden, A. & Nasibov, E. (2009). Nationalnyckeln tillSverigesfloraoch fauna. Tvåvingar: Blomflugor: Syrphinae. Diptera: Syrphidae: Syrphinae. *ArtDatabanken, SLU*, Uppsala, pp. 406.

Brădescu, V. (1975-1976). Considerații zoogeografice și ecologice privind fauna Sirfidelor (Diptera) din Parcul Național Retezat. *Sargetia, Acta Museii Devensis, Series Scientia Naturae 11-12,* 297-300.

Brădescu, V. (1986). Études diptérologique (Syrphidae) dans la réserve naturelle Domogled - Vallée de la Cerna. *Trav. Mus. Natl Hist. Nat. "Grigore Antipa" 28,* 121-131.

Brădescu, V. (1991). Les Syrphides de Roumanie (Diptera, Syrphidae). Clés de determination et repartition. *Trav. Mus. Natl. Hist. Nat. "Grigore Antipa" 31*, 7-83.

Brădescu, V. (1995). Données nouvelles concernant la faune diptérologique (Syrphidae) du Parc National Domogled - Vallée de la Cerna, Roumanie. *Trav. Mus. Natl. Hist. Nat. "Grigore Antipa" 35*, 419-422.

Burgio, G., Di Saverio, M., Crudele, G. & Sommaggio, D. (2025). Updated hoverfly (Diptera, Syrphidae) checklist of the Foreste Casentinesi National Park (Italy). *Biodivers. Data J. 13*, e147052.

Cicort-Lucaciu, A.-Ş., Bondar, A. & Covaciu-Marcov, S.-D. (2017). The alpine newt *Ichthyosaura alpestris* in Almăj Mountains, south-western Romania – 10 km upstream the Danube. *Oltenia. Studii şi Comunicări. Ştiinţele Naturii 33*, 1, 127-130.

Covaciu-Marcov, S.-D., Cicort-Lucaciu, A.-Ş., Gaceu, O., Sas, I., Ferenţi, S. & Bogdan, H. V. (2009). The herpetofauna of the south-western part of Mehedinţi County, Romania. *North-West. J. Zool. 5,* 1, 142-164.

Covaciu-Marcov, S.-D., Dumbravă, A.-R. & Ferenţi, S. (2025). Searching for the missing newts: notes on newt distribution in the Iron Gates Natural Park, Romania. *eco.mont 17,* 1, 17-24. https://doi.org/10.1553/eco.mont-17-1s17

Csürös, Ş., Pop, I., Hodişan, I. & Csürös-Káptlan, M. (1968). Cercetări floristice şi de vegetaţie între Orşova şi Eşelniţa. *Contribuţii Botanice 8*, 277-312.

de Groot, M. & Govedič, M. (2008). Checklist of the hoverflies (Diptera: Syrphidae) of Slovenia. *Acta Entomol. Slov. 16*, 1, 67-87.

Doyle, T., Hawkes, W.L.S., Massy, R., Powney, G.D., Menz, M.H.M. & Wotton, K.R. (2020). Pollination by hoverflies in the Anthropocene. *Proc. R. Soc. B 287*, 20200508. https://doi.org/10.1098/rspb.2020.0508

Dumbravă, A.-R. & Ruicănescu, A. (2023). Large, but hard to see: first record of a large European hoverfly *Milesia crabroniformis* (Diptera, Syrphidae) in Romania. *South-Western Journal of Horticulture, Biology and Environment 14*, 1, 51-56.

Dumbravă, A.-R. & Cupşa, D. (2025). Hoverflies (Diptera, Syrphidae) in the Danube Gorge, Carpathian Mountains, Romania: zoogeographical, ecological and conservation significance. *Acta Zool. Bulg. 77*, 4, 545-562. https://doi.org/10.71424/azb77.4.002930

Fijarczyk, A., Nadachowska, K., Hofman, S., Litvinchuk, S.N., Babik, W., Stuglik, M., Gollmann, G., Choleva, L., Cogălniceanu, D., Vukov, T., Džukić, G. & Szymura, J.M. (2011). Nuclear and mitochondrial phylogeography of the European fire-bellied toads *Bombina bombina* and *Bombina variegata* support their independent histories. *Mol. Ecol. 20,* 3381-3398. https://doi.org/10.1111/j.1365-294x.2011.05175.x

Finocchiaro, M., Médail, F., Saatkamp, A., Diadema, K., Pavon, D., Brousset, L. & Meineri, E. (2024). Microrefugia and microclimate: unraveling decoupling potential and resistance to heatwaves. *Sci. Total Environ. 924,* 171696. https://doi.org/10.1016/j.scitotenv.2024.171696

Gheorghe, M., Badea, G., Ene, T.E., Ioniţă, R.G. & Vîrghileanu, M.R. (2025). Geografie 8. *Editura Corint*, Bucureşti, pp. 128.

Haarto, A. & Kerppola, S. (2014). Checklist of the family Syrphidae (Diptera) of Finland. *ZooKeys 441,* 233-249. . https://doi.org/10.3897/zookeys.441.7251

Hågvar, E.B. & Nielsen, T.R. (2007). The hoverfly fauna (Diptera, Syrphidae) from six years of Malaise trapping in an organic barley field and its boundary in southern Norway. *Norw. J. Entomol. 54,* 135-145.

Haris, A., Józan, Z., Schmidt, P., Glemba, G., Tomozii, B., Csóka, G., Hirka, A., Šima, P. & Tóth, S. (2025). Climate change influences on central European insect fauna over the last 50 years: Mediterranean influx and non-native species. *Ecologies 6,* 16. https://doi.org/10.3390/ecologies6010016

Iacob, G.M., Manci, C.O., Craioveanu, C., Rákosy, L., Sitar, C. (2021). *Poecilocampa alpina* (Frey & Wullschlegel, 1874) (Lepidoptera, Lasiocampidae) first record from Romania. *Entomologica romanica 25,* 11-14.

Iftime, A. (2005). New observations on the herpetofauna from Domogled-Valea Cernei National Park and Porţile de Fier Natural Park (Romania). *Trav. Mus. Natl. Hist. Nat. "Grigore Antipa" 48*, 327-337.

Lebard, T. & Speight, M.C.D. (2019). Mise á jour de la liste des Syrphidae (Diptera) du Parc national du Mercantour. *Bull. Soc. Linn. Bordeaux 154*, 47, 1-2, 15-28.

Löhr, P.-W. (1990). Hoverflies (Diptera, Syrphidae) from Malaise traps in Ångermanland, coastal northern Sweden. *Entomol. Tidskr. 111*, 79-82.

Magyari, E.K., Jakab, G., Bálint, M., Kern, Z., Buczkó, K. & Braun, M. (2012). Rapid vegetation response to Lateglacial and early Holocene climatic fluctuations in the South Carpathian Mountains (Romania). *Quat. Sci. Rev. 35,* 116-130. https://doi.org/10.1016/j.quascirev.2012.01.006

Manci, C. & Dumbravă, A.-R. (2026). *Lindenia tetraphylla* (Odonata: Gomphidae) first record for Romania and future perspectives. *Entomologica romanica 30*, 1-4.

Mazánek, L., Hadrava, J., Semelbauer, M. & Král, M. (2025). Updated checklist of the Syrphidae Latreille, 1802 of Czechia and Slovakia. *Biodiversity and Environment 17 (special issue),* 127-141.

Milosavljević, M.J., Vujić, A., Popov, S., Radenković, S. & Miličić, M. (2026). Threatened or thriving? - the status of Serbian hoverflies on the IUCN European Red List of hoverflies. *Conserv. Sci. Pract. 8,* e70236. https://doi.org/10.1111/csp2.70236

Nielsen, T.R. & Svendsen, S. (2014). Hoverflies (Diptera, Syrphidae) in north Norway. *Norw. J. Entomol. 61*, 119-134.

Nilsson, S.G., Bygerjerg, R. & Franzén, M. (2007). Blomflugor (Diptera, Syrphidae) på en gård i Linnés hembygd i Stenbrohult. *Entomol. Tidskr. 128*, 4, 133-148.

Paşcovschi, S. (1956). Câteva consideraţii biogeografice asupra Munţilor Banatului. *Ocrotirea Naturii 2*, 111-134.

Paşcovschi, S. (1967). Succesiunea speciilor forestiere. *Editura Agro-Silvică*, Bucureşti, pp. 318.

Petruș-Vancea, A., Cupșa, D., Ferenți, S., Dumbravă, A.-R. & Covaciu-Marcov, S.-D. (2024). Low altitude *Vaccinium myrtillus* L. populations in the Eșelnița Valley (Danube Gorge, Carpathian Mountains, Romania). *J. Nat. Conserv. 82*, 126730. https://doi.org/10.1016/j.jnc.2024.126730

Reemer, M. (2009). De Nederlandese Zweefvliegen: Parasyrphus Roetneusjes. *Natuur van Nederland 8,* 271-275.

Reif, A., Schneider, E., Oprea, A., Rakosy, L. & Luick, R. (2022). Romania's natural forest types – a biogeographic and phytosociological overview in the context of politics and conservation. *Tuexenia 42,* 9-34. https://doi.org/10.14471/2022.42.005

Reverté, S., Miličić, M., Ačanski, J., Andrić, A., Aracil, A., Aubert, M., et al. (2023). National records of 3000 European bee and hoverfly species: a contribution to pollinator conservation. *Insect Conserv. Divers. 16*, 6, 758-775. https://doi.org/10.1111/icad.12680

Rösch, M. & Fischer, E. (2000). A radiocarbon dated Holocene pollen profile from the Banat Mountains (Southwestern Carpathians, Romania). *Flora 195*, 277-286. https://doi.org/10.1016/S0367-2530(17)30981-7

Ruicănescu, A. & Dumbravă, A.-R. (2020). First record of *Kisanthobia ariasi* (Coleoptera: Buprestidae) in Romania. *Trav. Mus. Natl. Hist. Nat. "Grigore Antipa" 63*, 2, 189-194. https://doi.org/10.3897/travaux.63.e56704

Sitar, C., Sielezniew, M., Malkiewicz, A., Frik, Z.F., Konvička, M. & Konvickova, H. (2025): Eastern arc of glacial relict species – population genetics of Violet Copper *Lycaena helle* butterfly in East-Central Europe. *Insects 16*, 1202. https://doi.org/10.3390/insects16121202

Schneider-Binder, E. (2014). Phytogeographical importance of the mountains along the Danube mountain gap valley and surrounding area. *Transylv. Rev. Syst. Ecol. Res. 16 – special issue, the "Iron Gates" Natural Park*, 11-28. https://doi.org/10.1515/trser-2015-0030

Schneider-Binder, E. (2016). Riparian vegetation on the left tributaries of the Danube along the "Clisura" cross valley. *Transylv. Rev. Syst. Ecol. Res. 18*, 3, 15-30. https://doi.org/10.1515/trser-2015-0091

Sentil, A., Miličić, M., Benrezkallah, J., et al. (2026). Synthesized database of wild bees and hoverfly records in Europe. *Sci. Data 13*, 227. https://doi.org/10.1038/s41597-026-06644-2

Słowińska, S., Ronikier, M., Paul, W., Kaszkiel, A., Kowalczyk, P. & Słowiński, M. (2026). The role of microclimate in supporting peatlands as climate-change refugia: a Central European perspective. *Conserv. Sci. Pract. 8,* e70172. https://doi.org/10.1111/csp2.70172

Speight, M.C.D. (2004). Towards an understanding of the development and constitution of the Irish postglacial syrphid fauna (Syrphidae, Diptera). *Volucella 7,* 125-155.

Speight, M.C.D. (2024). Species accounts of European Syrphidae, 2024. *Syrph the Net, the database of European Syrphidae (Diptera)*, *Syrph the Net publications,* Dublin, vol. 115, pp. 381.

Speight, M.C.D. & Sarthou, J.-P. (2017). StN keys for the species identification of the European species of various genera of Syrphidae, 2017/Clés StN pour la détermination des espèces Européennes de plusieurs genres des Syrphidae 2017. *Syrph the Net, the database of European Syrphidae (Diptera), Syrph the Net publications,* Dublin, vol. 99, pp. 139.

Stănescu, C. & Pârvu, C. (2005). Syrphids (Syrphidae, Diptera) of Romania, Checklist, Phenology, Distribution. *Trav. Mus. Natl. Hist. Nat. "Grigore Antipa" 48*, 177-202.

Stoenescu, Ş. M., Şchiop, A., Dica, I., Popescu, E., Patrichi, E. & Ţepeş, E. (1966). Atlasul climatologic al R. S. R. *Institutul Meteorologic, Bucureşti*.

Şuteu, D., Băcilă, I., Stoica, A.-I., Balász, Z.R., Puşcaş, M. & Coldea, G. (2023). Phylogeographic pattern of high alpine plant species *Eritrichium nanum* (Boraginaceae) within the Carpathians. *Not. Bot. Horti Agrobot. Cluj-Na. 51,* 1, 12971. https://doi.org/10.15835/nbha51112971

Takov, D., Georgieva, M., Ostoich, P., Pilarska, D. & Barta, M. (2025). Insects and their practical role in the functioning of human societies. *North-West. J. Zool. 21,* 2, 193-203.

Teodor, L.A., Ferenţi, S. & Covaciu-Marcov, S.D. (2019). Weevils Die in Vain? Understanding Messages from Road-Killed Weevils (Coleoptera: Curculionoidea). *Coleopt. Bull. 73*, 2, 359-368. https://doi.org/10.1649/0010-065X-73.2.359

Teodor, L.A., Lazăr, A.-M., Mușet, A.-F. & Dumbravă, A.-R. (2025). *Liparus glabrirostris* Küster, 1849 (Coleoptera: Curculionidae) in the Danube Gorge: a low altitude population in south-eastern Europe, zoogeographical and ecological consideration. *South-Western Journal of Horticulture, Biology and Environment 16*, 1, 70-79.

Tóth, S. (1995). Adatok a Mátra zengőlégy faunájához (Diptera: Syrphidae), II. Mátrakeresztes. *Folia Hist. Nat. Mus. Matr. 20*: 129-143.

Tóth, S. (2011). Magyarország zengőlégy faunája (Diptera: Syrphidae). *Acta Nat. Pannon., Supplementum 1*: 5-408.

Van de Meutter, F., Bot, S., Mortelmans, J. & Mengual, X. (2025). New hoverfly records for Greece with the description of a new species of *Cheilosia* (Diptera, Syrphidae). *Bol. Asoc. Esp. Entomol. 49*, 3-4, 201-225. https://doi.org/10.70186/baeeSGOU2274

van Steenis, J. (2011). Swedish hoverfly records (Diptera: Syrphidae). *Entomol. Tidskr. 132*, 3, 187-193.

van Steenis, W., de Groot, M. & van Steenis, J. (2013). New data on the hoverflies (Diptera: Syrphidae) of Slovenia. *Acta Entomol. Slov. 21*, 2, 131-162.

van Veen, M.P. (2004). Hoverflies of Northwest Europe: Identification keys to the Syrphidae. *KNNV Publishing*, pp. 254.

Verlinden, L. (1994). Faune de Belgique, Syrphides (Syrphidae). *Edition de l'Institut Royal des Sciences Naturelles de Belgique*, Bruxelles, pp. 289.

Voisin, C., Kirschner, P., Záveská, E., Frajman, B., Hülber, K., Wesseley, J., Willner, W., Schönswetter, P. & Carcinero, P. (2026). Spatiotemporal diversification of forest understorey species reveals the existence of multiple Pleistocene forest refugia in Central Europe. *Mol. Ecol. 35,* e70200. https://doi.org/10.1111/mec.70200

Vujić, A., Radenković, S.R., Nedeljković, Z.S. & Šimić, S. (2018). A new check list of hoverflies (Diptera: Syrphidae) of the Republic of Serbia. *Matica Srpska J. Nat. Sci. 135*, 7-51. https://doi.org/10.2298/ZMSPN1835007V

Zeegers, T., Van Steenis, W., Reemer, M. & Smit, J.T. (2024). Drastic acceleration of the extinction rate of hoverflies (Diptera: Syrphidae) in the Netherlands in recent decades, contrary to wild bees (Hymenoptera: Anthophila). *Journaal van Syrphidae 3*, 1, 1-11. https://doi.org/10.55710/1/YDSJ1547

Żoralski, R. & Kowalczyk, J.K. (2015). Syrphidae (Diptera) Trójmiejskiego Parku Krajobrazowego i terenów przyleglych. *Parki Narodowe i Rezerwaty Przyrody 34,* 1, 25-80.

Parasyrphus malinellus (Collin, 1952) in GBIF Secretariat (2023): GBIF Backbone Taxonomy. Checklist dataset https://doi.org/10.15468/39omei accessed via GBIF.org on 2026-01-15.

STUDIA UNIVERSITATIS BABEŞ-BOLYAI BIOLOGIA, 71, 1, 2026 (p. 17-38)
doi:10.24193/subbbiol.2026.1.02

Advances in molecular sexing of monomorphic birds in Indonesia: A scoping-review with potential application on the Bali starling (*Leucopsar rothschildi*) breeding

I Dewa Agung Panji Dwipayana[1]✉, Ni Luh Putu Kayika Febryanti[2], I Ketut Ginantra[1], Luh Putu Eswaryanti Kusuma Yuni[1,4], and Ni Luh Watiniasih[1,3]

[1]*Department of Biology, Faculty of Mathematics and Natural Sciences, Universitas Udayana, Bali, Indonesia;* [2]*Department of Aquaculture, Faculty of Marine Science and Fisheries, Universitas Udayana, Bali, Indonesia;* [3]*Animal Taxonomy Laboratory, Universitas Udayana, Bali, Indonesia;* [4]*Ornithology Research Center, Universitas Udayana, Bali, Indonesia*
✉***Corresponding author, E-mail: panji.dwipayana@unud.ac.id***

Article history: Received 20 March 2026; Revised 18 May 2026; Accepted 18 May 2026; Available online 25 June 2026

Abstract. Accurate sex determination is essential for ecological research, captive breeding, and conservation of sexually monomorphic birds. In Indonesia, despite exceptional avian diversity and urgent conservation needs, the application of molecular sexing remains uneven across taxa. This scoping review synthesizes evidence from 35 studies published between 2015 and 2025, drawn from 176 records, that employed PCR amplification of chromo-helicase-DNA-binding (CHD) genes using various primer sets and sample types. Blood and feathers were the most commonly used materials (≈80% of records), while swabs, eggshell membranes, and casque tissue were reported infrequently. Primer performance varied markedly among avian families and sample types. CHD1LF/CHD1LR and NP + P2 + MP consistently showed high amplification reliability, whereas P2/P8 and 1237L/1272H yielded inconsistent results. To enable objective comparison, we propose the Primer Performance Index (PPI), a composite metric integrating amplification success with

taxonomic breadth. PPI analysis demonstrated that primers with broad family coverage maintained high reliability after normalization, while taxa-restricted primers showed reduced performance. These findings support improved primer selection for conservation breeding, particularly for the critically endangered Bali Starling (*Leucopsar rothschildi*), and promote more standardized integration of molecular sexing into avian conservation practice in Indonesia.

Keywords: Bali Starling, conservation genetics, molecular sexing, Primer Performance Index.

Introduction

Accurate sex determination in birds is a fundamental requirement for ecological research, captive breeding, and conservation management. In sexually monomorphic species, where males and females lack obvious external differences, traditional approaches such as behavioral observation or morphometrics are unreliable, particularly in juveniles (Petrou *et al.,* 2024). Gender misidentification can lead to skewed datasets in ecological studies and failed breeding programs (Hall *et al.,* 2025).

Molecular sexing, based on amplification of introns within the chromo-helicase-DNA-binding (CHD) genes on the Z and W chromosomes, offers a powerful solution. A range of primer sets has been developed to target these regions, including P2/P8, 2550F/2718R, CHD1F/CHD1R, CHD1LF/CHD1LR, and NP + P2 + MP. Reported success rates vary across taxa and sample types, suggesting that primer performance is context-dependent (Fitriana *et al.,* 2022; Ratri *et al.,* 2022; Fitriana *et al.,* 2023b; Fitriana *et al.,* 2023c). Blood is often considered the most reliable source of DNA, while feathers provide a non-invasive alternative but are prone to degradation. Other tissues such as swabs and eggshell membranes have been tested sporadically, but their limited representation raises questions about generalizability (Akrom *et al.,* 2020b; Yuda *et al.,* 2020; Vera *et al.,* 2021; Pamulang and Haryanto, 2021).

Indonesia has extraordinary avian diversity and pressing conservation challenges. It provides a critical context for evaluating molecular sexing. Despite the country's high bird richness and the urgent need for effective breeding management in threatened species, molecular sexing has been applied unevenly, with fragmented coverage across families and limited integration into conservation practice (Purwaningrum *et al.,* 2019, Ratri *et al.,* 2022). The critically endangered Bali Starling (*Leucopsar rothschildi*) exemplifies this gap (IUCN, 2021). Although extensive rehabilitation and release efforts have been undertaken (Sutomo *et al.,* 2023), molecular sexing has rarely been incorporated into breeding programs.

This review seeks to synthesize current evidence on molecular sexing of monomorphic birds in Indonesia, with particular emphasis on blood and feather samples as the most extensively tested and practical materials. It aims to identify patterns in how different sample types and primer sets influence amplification success across avian families. A new framework named the Primer Performance Index (PPI), is introduced to normalize success rates and provide a more balanced comparison of primer reliability. The resulting insights are then applied to explore opportunities for strengthening the breeding management of the critically endangered Bali starling (*Leucopsar rothschildi*).

Materials and methods

Methods

This scoping review followed the methodological and reporting standards outlined in the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) framework as established (Tricco *et al.,* 2018). The entire document, including the flow diagram presented in Figure 1, was structured to comply with all 22 items specified in the PRISMA-ScR checklist.

Protocol and Registration

All methodological procedures and decisions are comprehensively detailed within the body of this manuscript.

Inclusion Criteria

This scoping review applied a set of predefined inclusion and exclusion criteria to identify relevant studies on molecular sexing of monomorphic birds in Indonesia, with particular emphasis on the Bali starling (*Leucopsar rothschildi*). Studies were first de-duplicated using Zotero, then included if they met all of the following conditions: (1) they were accessible primary research articles, published in peer-reviewed journals or institutional repositories; (2) they focused on the sexing of bird species found in Indonesia, including but not limited to the Bali starling; (3) they involved monomorphic birds, which are species lacking external sexual dimorphism; (4) they were written in English or Bahasa Indonesia and published between January 2015 and October 2025; and (5) they reported at least one technical detail such as sample type, gene target, primer identity, PCR conditions, or sexing accuracy.

Exclusion Criteria

Studies were excluded if they focused on sexually dimorphic birds or non-avian taxa, relied solely on morphological, behavioral, or hormonal sexing methods without molecular data, or lacked sufficient methodological detail. Review articles, opinion pieces, conference abstracts, and theses without full-text access or complete workflows were also excluded. Additionally, studies conducted entirely outside Indonesia with no relevance to Indonesian bird species or populations were not considered.

Information Sources

A comprehensive literature search was carried out in November 2025 across four electronic databases: Google Scholar, PubMed, the Directory of Open Access Journals (DOAJ), and GARUDA, an Indonesian journal indexing platform. To supplement the database search, reference lists from key publications were manually examined to identify additional supportive studies.

Search Strategy

To maintain relevance and reduce retrieval of low-quality or random results, the search was restricted to the first ten pages of each keyword query across all databases. The complete search strings are documented in Supplementary Material Table S1 in (https://osf.io/hwbru/files/osfstorage/69b206346d96b8672c0514e9).

Example for Google Scholar: "Bird" AND ("Indonesia" OR “Bali”) AND ("molecular sexing" OR "PCR sexing") AND ("Primer").

Selection of Sources of Evidence

The titles and abstracts of all 176 records were independently screened by the first and second authors in accordance with the predefined eligibility criteria. Discrepancies identified during either the initial screening or the full-text assessment were addressed through discussion and, when required, resolved collectively by all authors. The third author subsequently verified and finalized the extracted metadata against the original publications to ensure accuracy and consistency. The complete study selection process is illustrated in the PRISMA-ScR flow diagram shown in Figure 1.

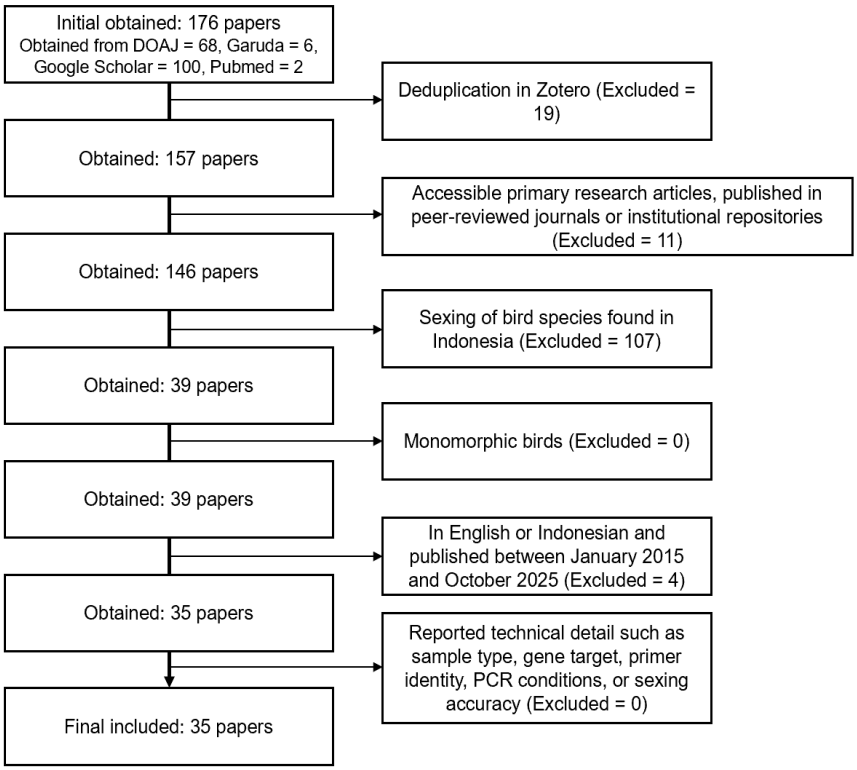


Figure 1. PRISMA-ScR flow diagram. The initial papers obtained were 176, the final included papers were 35.

Data Charting Process

A standardized data-charting template was created in Excel and initially piloted on five journals to ensure consistency. Following this calibration, data extraction was carried out independently by the first and second authors, with any discrepancies resolved through consensus discussions involving the third author.

Data Items

For each included study, we extracted bibliographic details (authors, year, journal name), species, family, and geographic origin. Molecular sexing details included sample type (blood, feather, swabs), DNA extraction method/kit, primer name, primer sequence, gene target name, PCR conditions (e.g. annealing temperature, cycle number), and gel electrophoresis resolution. We recorded sexing success rate, false positives/negatives, and comparison method (e.g., against behavioral method).

Synthesis of Result

All primers from the collected studies were consolidated and plotted against the families in which they were applied. When the same primer appeared in multiple studies targeting the same family, its success rates were averaged to produce a single representative value for that primer-family combination. For each primer within each sample type category (blood, feathers, oral swabs, anal swabs, casque tissue), two values were calculated: the average success rate across individuals, and the number of families represented. These values were used to generate a standardized measure of primer efficiency through the Primer Performance Index (PPI), which integrates raw success with taxonomic breadth. By applying this procedure separately for each sample type, the synthesis produced a structured dataset in which primer performance could be compared across tissues and families on a normalized basis.

In this index, primer success rate is weighted by the proportion of avian families in which the primer was tested. This adjustment is necessary because a primer that shows 100% success in only one family should not be interpreted as equally reliable as a primer that maintains high success across many families. Therefore, PPI gives higher comparative value to primer sets that combine high amplification success with broader taxonomic coverage.
Primer Performance Index (PPI) is calculated by:

$$PPI = (SR\ x\ 0.5) + (\frac{NF}{TF}\ x\ 0.5)$$

where:

- SR = Primer success rate for a certain sample across families
- NF = Number of families tested by this primer
- TF = Total number of families included in this study

Results

A complete extracted data is available in the Supplementary Material Table S2 (https://osf.io/hwbru/files/osfstorage/69b206346d96b8672c0514e9).

Selection of Sources of Evidence

All titles and abstracts were independently assessed by two reviewers in accordance with the predefined eligibility criteria outlined in the Method section.

Characteristics of Sources of Evidence

Family distribution from the obtained studies were calculated for each species included. Sample type distribution is calculated by multiplying each species with number of sample types taken from them. Thus, the number will be higher than the total species included because one individual could have multiple sample types (e.g. blood, feather) taken from them.

The dataset in Figure 2 includes molecular sexing records from 26 avian families and 67 species. *Accipitridae* accounted for the largest proportion of samples (16%), followed by *Cacatuidae* (14%). *Columbidae, Psittaculidae*, and *Psittacidae* represented 9% each, while *Phasianidae* contributed 7%. The remaining families, each only accounted for < 5% of the dataset. This distribution is not perfectly equal, but it is sufficiently balanced to allow robust amplification trends to be identified in well-sampled groups, while still capturing comparative insights across many others.

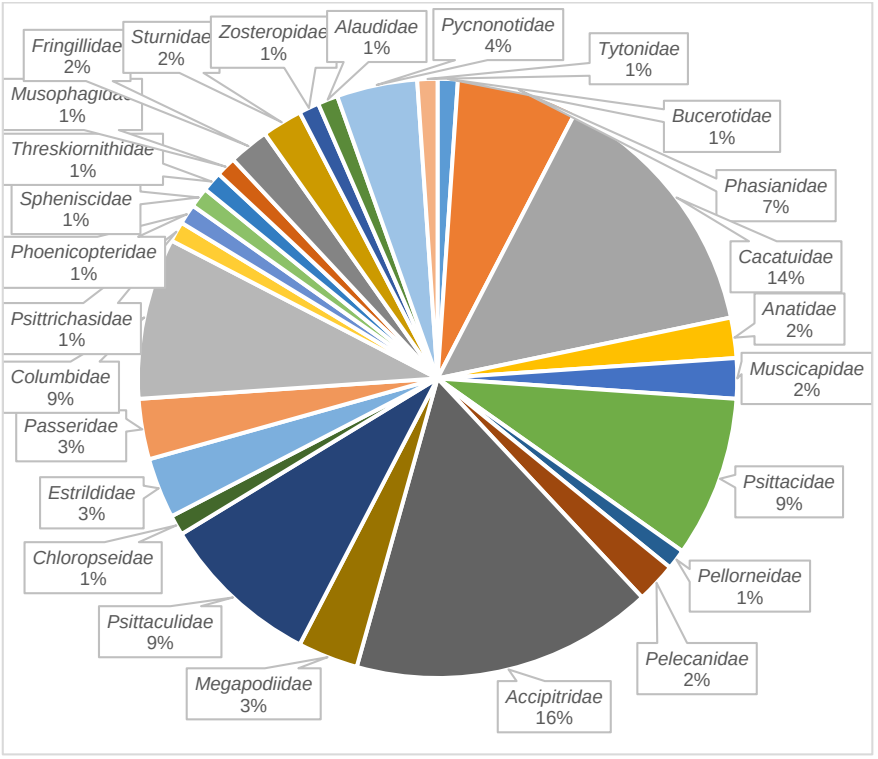


Figure 2. Family-level representation of birds examined in molecular sexing studies included in this review. Percentages indicate the proportion of extracted family records across the included studies. Studies involving more than one family were counted separately for each family

A total of six biological sample types were used for molecular sexing, with blood samples comprising the majority (50%) of the dataset (Figure 3). Feathers accounted for 30%, reflecting their practical use in non-invasive sampling. Swab-based methods contributed a combined 17%, split between oral swabs (9%) and anal swabs (8%), while eggshell membranes (2%) and casque tissue (1%) were used in specialized cases. This distribution highlights a strong reliance on blood and feather samples for routine sexing, while also demonstrating the feasibility of alternative sources in forensic or conservation contexts.

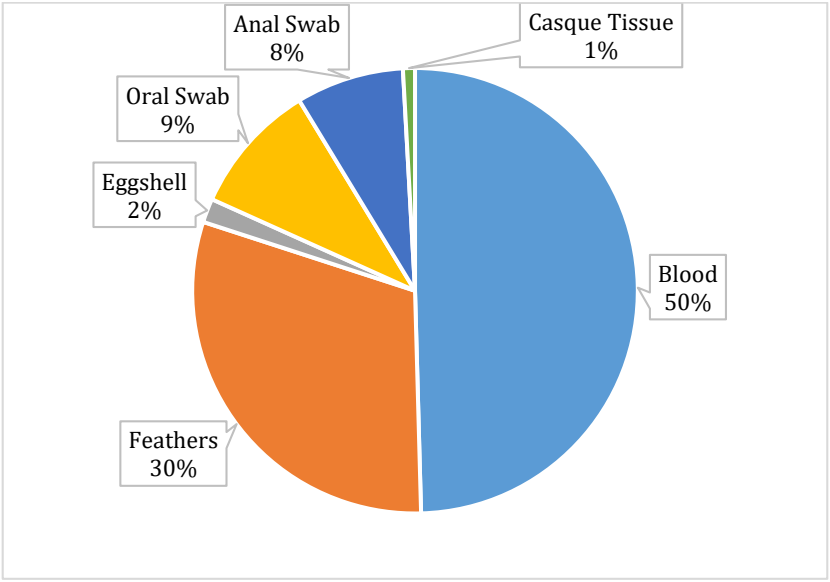


Figure 3. Distribution of biological sample types used for molecular sexing in the studies included in this scoping review. Percentages indicate the proportion of extracted sample-type records across the included studies. When a study used more than one sample type, each sample type was counted separately

Results of Individual Sources of Evidence

Among the evaluated primer sets, the CHD1LF/CHD1LR primer pair achieved 100% PCR-based molecular sexing success rates across seven families (*Phasianidae, Anatidae, Muscicapidae, Chloropseidae, Psittacidae, Estrildidae,* and *Passeridae*) and maintaining values above 90% in blood samples overall (Table 1). CHD1F/CHD1R primer set also performed strongly, with complete success in *Sturnidae, Phasianidae, Muscicapidae*, and *Passeridae*, though values

dropped to 40% in *Columbidae*. The widely used 2550F/2718R primer set achieved 100% success in *Accipitridae, Megapodiidae*, and *Psittaculidae*, but showed reduced amplification in *Columbidae* (40%) and mixed outcomes in *Muscicapidae* (50%). NP + P2 + MP primer set was particularly effective in captive birds such as parrots, lovebirds, and cockatoos, reaching 100% in multiple *Psittacidae, Psittaculidae*, and *Cacatuidae* species, though feather success rates varied from 20% to 100%. InSex-F/InSex-R primer set achieved 100% in *Accipitridae* and *Phasianidae*, but dropped to 84.6% in degraded eggshell samples. In contrast, 1237L/1272H and CHDZ/CHDW primer sets frequently failed, with success rates mostly ranging from 0-50% depending on family, and P2/P8 primer set produced inconsistent values, from 0% in several *Psittacidae* species to 100% in, for example, *Zosteropidae, Alaudidae, Psittaculidae, Cacatuidae*, and *Pellorneidae*.

A complete list of extracted data is available in Supplementary Material Table S2, but a snip is presented below.

Table 1. Molecular sexing success rate according to species, family, sample type, and primer set in the included studies.

Title	Species	Family	Sample	Sexing Success Rate
(Fitriana *et al.,* 2020)	*Rhinoplax vigil*	*Bucerotidae*	Casque tissue	- CHD1F/CHD1R: 80.88% - 2550F/2718R: 51.47% - P2/P8: 20.58%
	Leucopsar rothschildi	*Sturnidae*	- Blood - Feathers	CHD1F/CHD1R: - Blood: 100% - Feather: 100% 2550F/2718R: - Blood: 100% - Feathers: 100% P2/P8: - Blood: 40.9% - Feathers: 0%
(Pambuko *et al.,* 2023)	*Gallus gallus domesticus*	*Phasianidae*	Blood	- CHD1F/CHD1R: 100% - AV2550F/AV2718R: 100% - AVP2/AVP8: 60% - AV1237L/AV1272H: 60% - CHDZ/CHDW: 0%
(Zein *et al.,* 2017)	*Cacatua galerita*	*Cacatuidae*	Blood	- 2550F/2718R: 100%
	Cacatua sulphurea	*Cacatuidae*		
	Cacatua alba	*Cacatuidae*		
	Cacatua moluccensis	*Cacatuidae*		

Title	Species	Family	Sample	Sexing Success Rate
	Cacatua goffiniana	*Cacatuidae*		
	Probosciger aterrimus	*Cacatuidae*		
(Fitriana *et al.,* 2023a)	*Gallus gallus domesticus*	*Phasianidae*	Blood	- CHD1LF/CHD1LR: 100% - CHD1F/CHD1R: 100% - 2550F/2718R: 100% - 1237L/1272H: 0% - P2/P8: 0%
	Anas platyrhynchos domesticus	*Anatidae*		-CHD1LF/CHD1LR: 100% - CHD1F/CHD1R: 100% - 2550F/2718R: 50% - 1237L/1272H: 0% - P2/P8: 0%
	Brachypteryx leucophrys	*Muscicapidae*		- CHD1LF/CHD1LR: 100% - CHD1F/CHD1R: 100% - 2550F/2718R: 50% - 1237L/1272H: 0% - P2/P8: 0%
	Agapornis fischeri	*Psittacidae*		- CHD1LF/CHD1LR: 100% - CHD1F/CHD1R: 0% - 2550F/2718R: 100% - 1237L/1272H: 0% - P2/P8: 0%
(Ó Marcaigh *et al.,* 2021)	*Pellorneum celebense*	*Pellorneidae*	Feathers	- P2/P8: 100% - CHD1F/CHD1R: 0%
(Dharmayanthi *et al.,* 2021)	*Pelecanus conspicillatus*	*Pelecanidae*	Blood	- 2550F/2718R: 100%
(Annisa *et al.,* 2021)	*Nisaetus cirrhatus*	*Accipitridae*	Blood	- 2550F/2718R: 100%
(Yuda and Saputra, 2021)	*Macrocephalon maleo*	*Megapodiidae*	Eggshell membrane	- 2550F/2718R: 100%
(Nugroho and Zein, 2015)	*Lorius lory*	*Psittaculidae*	Feathers	- 2550F/2718R: 100%
(Fuadah *et al.,* 2025)	*Copsychus malabaricus*	*Muscicapidae*	Blood	- CHD1LF/CHD1LR: 100% - CHD1F/CHD1R: 100% - P2/P8: 100% - 2550F/2718R: 50% - 1237L/1272H: 50%

Title	Species	Family	Sample	Sexing Success Rate
	Chloropsis sonnerati	*Chloropseidae*		- CHD1LF/CHD1LR: 100% - CHD1F/CHD1R: 50% - P2/P8: 0% - 2550F/2718R: 50% - 1237L/1272H: 0%
(Fitriana *et al.,* 2023b)	*Coturnix coturnix*	*Phasianidae*	Blood	CHD1LF/CHD1LR: 100%
	Melopsittacus undulatus	*Psittacidae*		
	Estrildid finches	*Estrildidae*		
	Passer montanus	*Passeridae*		
(Fitriana *et al.,* 2023c)	*Coturnix coturnix*	*Phasianidae*	Blood	- CHD1F/CHD1R: 100% - 2550F/2718R: 100% - 1237L/1272H: 50% - P2/P8: 50%
	Estrildid finches	*Estrildidae*		- CHD1F/CHD1R: 100% - 2550F/2718R: 50% - 1237L/1272H: 0% - P2/P8: 0%
	Passer montanus	*Passeridae*		- CHD1F/CHD1R: 100% - 2550F/2718R: 50% - 1237L/1272H: 0% - P2/P8: 0%
	Melopsittacus undulatus	*Psittacidae*		- CHD1F/CHD1R: 100% - 2550F/2718R: 100% - 1237L/1272H: 0% - P2/P8: 0%
(Fitriana, *et al.,* 2022)	*Columba livia*	*Columbidae*	Blood	- CHD1LF/CHD1LR: 90% - P2/P8: 20% - 2550F/2718R: 40% - CHD1F/CHD1R: 40% - 1237L/1272H: 50%
	Streptopelia chinensis	*Columbidae*		
	Treron griseicauda	*Columbidae*		
	Spilopelia chinensis	*Columbidae*		
	Geopelia striata	*Columbidae*		
(Ratri *et al.,* 2022)	*Cacatua sulphurea*	*Cacatuidae*	- Blood - Feathers	NP + P2 + MP (combined): - Blood: 100% - Feathers: 25%
(Savitri *et al.,* 2021)	*Cacatua galerita*	*Cacatuidae*	- Blood - Feathers	NP + P2 + MP (combined): - Blood: 100% - Feathers: 100%
(Hidayat *et al.,* 2021)	*Cacatua goffiniana*	*Cacatuidae*	- Blood - Feathers	NP + P2 + MP (combined): - Blood: 100% - Feathers: 20%
(Argarini *et al.,* 2020)	*Agapornis fischeri*	*Psittaculidae*	Feathers	NP + P2 + MP (combined): 100%

Title	Species	Family	Sample	Sexing Success Rate
(El Islami *et al.*, 2021)	*Agapornis personata*	*Psittaculidae*	Feathers	NP + P2 + MP (combined): 100%
(Sitohang, 2017)	*Nisaetus bartelsi*	*Accipitridae*	Blood	- 2550F/2718R: 100% - 2561/2728: 100% - P2/P8: 0% - 1237L/1272H: 0%
	Nisaetus cirrhatus	*Accipitridae*		
(Purwaningrum *et al.*, 2019)	*Eclectus roratus*	*Psittaculidae*	Feathers	NP + P2 + MP (combined): 100%
	Cacatua sanguinea	*Cacatuidae*		
	Probosciger aterrimus	*Cacatuidae*		
	Lorius lory	*Psittaculidae*		
	Psittrichas fulgidus	*Psittrichasidae*		
	Pelecanus onocrotalus	*Pelecanidae*		
	Cygnus atratus	*Anatidae*		
	Phoenicoparrus minor	*Phoenicopteridae*		
	Ara ararauna	*Psittacidae*		
	Eolophus roseicapilla	*Cacatuidae*		
	Spheniscus demersus	*Spheniscidae*		
	Psittacus erithacus	*Psittacidae*		
	Amazona amazonica	*Psittacidae*		
	Aratinga solstitialis	*Psittacidae*		
	Eudocimus ruber	*Threskiornithidae*		
	Tauraco persa	*Musophagidae*		
(Withaningsih *et al.*, 2024)	*Nisaetus bartelsi*	*Accipitridae*	Blood	- 2550F/2718R: 100%
(Akrom *et al.*, 2020b)	*Serinus canaria*	*Fringillidae*	- Blood - Feathers	CHD1F/CHD1R: - Adult males blood and feather: 100%. - Adult females blood and feather: 100%. - Juveniles feather: 100%
(Samad *et al.*, 2023)	*Macrocephalon maleo*	*Megapodiidae*	- Feathers - Eggshell	2550F/2718R: - Feather: 100% - Eggshell: 100%.
(Adelia *et al.*, 2025)	*Columba livia*	*Columbidae*	Feathers	2550F/2718R: 100%
(Wirastika *et al.*, 2015)	*Leucopsar rothschildi*	*Sturnidae*	Feathers	- P2/P8: 90% - 2550F/2718R: 86.7% - 1237L/1272H: 73.3%

Title	Species	Family	Sample	Sexing Success Rate
(Nugraheni *et al.*, 2019)	*Agapornis roseicollis*	*Psittaculidae*	Feathers	NP + P2 + MP (combined): 100%
(Azzahra *et al.*, 2025)	*Columba livia*	*Columbidae*	Blood	2550F/2718R: 100%
(Oktanella, 2017)	*Zosterops flavus*	*Zosteropidae*	Feather	- P2/P8: 100%
	Mirafra javanica	*Alaudidae*		
	Agapornis sp.	*Psittaculidae*		
	Nymphicus hollandicus	*Cacatuidae*		
(Pinayungan *et al.*, 2024)	*Aratinga solstitialis*	*Psittacidae*	Feather	P2/P8: 100%
(Fatona *et al.*, 2025)	*Agapornis fischeri*	*Psittaculidae*	Feathers	CHD1F/CHD1R: 100%
(Vera *et al.*, 2021)	*Nisaetus cirrhatus*	*Accipitridae*	Blood	- InSex-F/InSex-R: 100%
	Haliastur indus	*Accipitridae*	Blood	
	Passer montanus	*Passeridae*	Blood	
	Lonchura fuscans	*Estrildidae*	Blood	
	Pycnonotus zeylanicus	*Pycnonotidae*	Blood	
	Tyto alba	*Tytonidae*	Feather	
	Gallus gallus	*Phasianidae*	Blood	
	Pavo muticus	*Phasianidae*	Blood	
	Macrocephalon maleo	*Megapodiidae*	Eggshell membrane	- InSex-F/InSex-R: 84.6%; 2 failed samples likely due to low quality from eggshell
(Pamulang and Haryanto, 2021)	*Pycnonotus aurigaster*	*Pycnonotidae*	- Blood - Feathers	P2/P8: - Blood: 100% - Feather: 50%
	Pycnonotus melanicterus	*Pycnonotidae*		
(Akrom *et al.*, 2020a)	*Serinus canaria*	*Fringillidae*	Buccal swab	CHD1F/CHD1R: - Adults: 100% - Nestlings: 100%
	Columba livia (Pigeon)	*Columbidae*		
(Yuda *et al.*, 2020)	*Nisaetus cirrhatus*	*Accipitridae*	- Blood - Tracheal swab - Cloacal swab	2550F/2718R: - Blood: 100% - Tracheal swab: 71% - Cloacal swab: 9%
	Spilornis cheela	*Accipitridae*		
	Nisaetus bartelsi	*Accipitridae*		
	Ichthyophaga leucogaster	*Accipitridae*		
	Ichthyophaga ichtyaetus	*Accipitridae*		

Title	Species	Family	Sample	Sexing Success Rate
	Ictinaetus malaiensis	*Accipitridae*		
	Milvus migrans	*Accipitridae*		
	Tachyspiza virgata	*Accipitridae*		
	Pernis ptilorhyncus	*Accipitridae*		
(Yuda and Wajjwalku, 2022)	*Pycnonotus zeylanicus*	*Pycnonotidae*	Feathers	- 2561-w/2728-w: 100%

Although eggshell, oral swab, anal swab, and casque data are reported here for completeness, each represent fewer than 10% of all records and will not be examined in detail in the discussion. Similarly, family-level comparisons are not pursued further because 14 families were represented by only one or two studies. Emphasis in the discussion is placed on sample type and primer performance, which provide more robust and generalizable insights into molecular sexing accuracy.

Synthesis of Primer Success Rate and Performance Index Results

Overall, the synthesis revealed that blood samples consistently produced the highest success rates across primer sets, while feathers yielded more variable outcomes and other types showed limited reliability. Primers such as CHD1LF/CHD1LR and NP + P2 + MP primer sets retained strong performance even after normalization, confirming their broad applicability, whereas set like 2561-w/2728-w showed reduced values once PPI was applied, reflecting their limited generalizability. By combining raw success rates with PPI, the analysis provided a balanced view of primer efficiency, highlighting those with reproducible utility across families and sample types while controlling for bias introduced by uneven representation

Primer Performance Index Calculation

Analysis shown here was restricted to blood and feathers samples, as eggshell, oral swabs, anal swabs, and casque tissue were excluded due to the limited number of records available. For each primer, weighed success rates were calculated alongside the number of families. These values were used to derive the Primer Performance Index (PPI) in Table 2.

Table 2. Primer Performance Index Calculation

Blood			Feathers		
Primer Type	**Average Primer Success Rate**	**Primer Performance Index**	**Primer Type**	**Average Primer Success Rate**	**Primer Performance Index**
1237L/1272H	11.1%	0.13	1237L/1272H	73.3%	0.38
2550F/2718R	75.0%	0.47	2550F/2718R	98.3%	0.56
2561/2728	100.0%	0.51	2561/2728	-	-
InSex-F/InSex-R	100.0%	0.54	InSex-F/InSex-R	100.0%	0.52
P2/P8	21.4%	0.20	P2/P8	86.9%	0.57
CHD1F/CHD1R	84.0%	0.50	CHD1F/CHD1R	75.0%	0.44
CHD1LF/CHD1LR	98.8%	0.56	CHD1LF/CHD1LR	-	-
NP + P2 + MP	100.0%	0.51	NP + P2 + MP	97.4%	0.66
AV1237L/AV1272H	60.0%	0.31	AV1237L/AV1272H	-	-
AV2550F/AV2718R	100.0%	0.51	AV2550F/AV2718R	-	-
AVP2/AVP8	60.0%	0.31	AVP2/AVP8	-	-
CHDZ/CHDW	0.0%	0.01	CHDZ/CHDW	-	-
2561-w/2728-w (Direct PCR)	-	-	2561-w/2728-w (Direct PCR)	100.0%	0.52

In blood samples, the highest primer performance was observed for CHD1LF/CHD1LR primer set (PPI 0.56, 98.8% success), followed closely by InSex-F/InSex-R primer set (0.54, 100%) and CHD1F/CHD1R primer set (0.50, 84%). Other primers such as NP + P2 + MP, AV2550F/AV2718R, and 2561/2728 primer sets maintained consistent efficiency (PPI 0.51, 100%). In contrast, P2/P8 (0.20, 21.4%), AVP2/AVP8 (0.31, 60%), AV1237L/AV1272H (0.31, 60%), 1237L/1272H (0.13, 11.1%), and CHDZ/CHDW (0.01, 0%) primer sets performed poorly. In feather samples, NP + P2 + MP primer set achieved the strongest outcome (PPI 0.66, 97.4%), followed by P2/P8 (0.57, 86.9%) and 2550F/2718R (0.56, 98.3%) primer sets. InSex-F/InSex-R and 2561-w/2728-w primer sets (Direct PCR) both reached solid values (0.52, 100%), while CHD1F/CHD1R primer set showed moderate efficiency (0.44, 75%) and 1237L/1272H primer set remained the weakest (0.38, 73.3%). These results provide a clear comparative overview of primer reliability across sample types, reinforcing the value of high-performing primers for consistent molecular sexing.

Discussion

Influence of Sample Type

Blood samples comprise 50% of records. Among all biological materials used for avian molecular sexing, blood remains the most consistently reliable source of DNA. It shows near-universal amplification success across taxa, primer sets, and laboratory conditions, with exception in low performing primers such as 1237L/1272H primer set (Fitriana *et al.,* 2023a). Studies covering both captive and wild birds repeatedly report clear CHD-band separation with minimal PCR failure (Zein *et al.,* 2017; Yuda *et al.,* 2020). The presence of nucleated erythrocytes in birds allows DNA extracted from blood to retain high integrity, producing strong template yield with little need for repeat sampling or nested PCR (Turcu *et al.,* 2023). For these reasons, blood continues to be regarded as the benchmark sample type for molecular sex determination in birds. However, its invasive nature raises ethical and logistical concerns. Repeated venipuncture in endangered species and juveniles can cause stress, requires trained personnel, and is difficult to scale in field conditions (Marc *et al.,* 2025).

Feathers, representing 30% of records, illustrate the trade-off between accessibility and reliability. Success ranged from 0% (CHD1F/CHD1R primer set in *Agapornis fischeri* (Fitriana *et al.,* 2023a) to 100% (NP + P2 + MP primer set in *Psittaculidae* and *Cacatuidae* (Purwaningrum *et al.,* 2019; Nugraheni *et al.,* 2019). DNA extracted from *Columbidae* feathers were amplified at 100% with 2550F/2718R primer set but only 40% with CHD1F/CHD1R and 20% with P2/P8 primer sets (Fitriana *et al.,* 2022; Azzahra *et al.,* 2025). This heterogeneity underscores that molecular sexing analysis based on DNA extracted from feathers is not inherently unreliable but contingent on primer choice and DNA integrity. DNA from degraded calamus tissue often reduces amplification, yet when paired with robust primers such as NP + P2 + MP primer set with overnight lysis buffer incubation, feathers can rival blood in accuracy (Argarini *et al.,* 2020; El Islami *et al.,* 2021). Direct PCR is also a great boost for feather samples. It achieved 100% success in *Pycnonotidae* (Yuda and Wajjwalku, 2022). This means feathers are a viable substitute if protocols explicitly address DNA degradation risk by standardizing feather type, storage, and extraction methods.

Influence of Represented Families, Primer Type, and Primer Performance Index

Figure 2 confirms that raptors (*Accipitridae*, 16%) and cockatoos (*Cacatuidae*, 14%) dominate Indonesian molecular sexing research, followed by *Columbidae, Psittaculidae*, and *Psittacidae* (each 9%), and *Phasianidae* (7%). This concentration

reflects both conservation priorities and socio-economic drivers: raptors are frequently confiscated and require sexing for rehabilitation, while cockatoos and parrots are central to aviculture and trade (Zein *et al.,* 2017; Yuda *et al.,* 2020). The consequence is methodological bias. Primers validated in these families, such as 2550F/2718R and NP + P2 + MP, appear favorable, but their reliability is partly a product of repeated testing in taxa with suitable CHD intron structures (Yuda *et al.,* 2020; Savitri *et al.,* 2021; Hidayat *et al.,* 2021; Ratri et al., 2022). To address this imbalance, we propose the Primer Performance Index (PPI) as a descriptive normalization metric that combines primer success rate with the breadth of taxonomic testing. The index is intended to reduce overinterpretation of high success rates obtained from narrowly tested primers and to provide a more balanced comparison across primer sets.

Primer performance differed markedly between blood and feathers. CHD1LF/CHD1LR primer set was the most reliable, achieving 100% success in DNA extracted from blood across *Phasianidae, Anatidae, Muscicapidae, Chloropseidae, Psittacidae, Estrildidae*, and *Passeridae*, and maintaining values above 90% overall (Fitriana *et al.,* 2022; Fitriana *et al.,* 2023a; Fitriana *et al.,* 2023b; Fuadah *et al.,* 2025). NP + P2 + MP primer set achieved 100% in DNA extracted from feathers across *Psittaculidae, Cacatuidae, Psittrichasidae, Pelecanidae, Anatidae, Phoenicopteridae, Psittacidae, Spheniscidae, Threskiornithidae,* and *Musophagidae,* though values dropped to 20-25% in some cockatoo species (Purwaningrum *et al.,* 2019; Nugraheni *et al.,* 2019; Hidayat *et al.,* 2021, Ratri *et al.,* 2022). By contrast, P2/P8 was highly inconsistent. It achieved 0% in *Psittacidae* and 100% in *Pellorneidae* and *Zosteropidae* (Oktanella, 2017; Ó Marcaigh *et al*., 2021; Fitriana *et al.,* 2023c). The performance of 1237L/1272H primer set remained weak, with mostly between 0-50% depending on family (Wirastika *et al.,* 2015; Fitriana *et al.,* 2023c).

The Primer Performance Index (PPI) confirmed these trends. CHD1LF/ CHD1LR and NP + P2 + MP primer sets retained strong values after normalization, underscoring their robustness. 2561-w/2728-w showed inflated raw success in DNA extracted from feathers (up to 100%) but dropped after normalization, revealing limited generalizability. 2561-w/2728-w primer set's initial apparent efficiency was inflated by success in limited taxa, while PPI corrected this inflation by taking into account its narrow taxa. 1237L/1272H primer set performance remained consistently poor, confirming its restricted applicability.

These results highlight the importance of considering PPI validation, not just raw primer efficiency. A primer tested across many families with moderate success may be more reliable in practice than one showing perfect success in a single species (Marc *et al.,* 2025). Raw success for context or taxa specific application is a strength, but PPI is essential for cross-taxa analysis to avoid overinterpretation.

Implications for the Bali Starling (Leucopsar rothschildi) Breeding

Two studies provide complementary insights into primer performance. The 2015 study reported sexing success rates of 90% with P2/P8, 86.7% with 2550F/2718R, and 73.3% with 1237L/1272H primer sets. Meanwhile, a later study demonstrated 100% success with CHD1F/CHD1R and 2550F/2718R primer sets in both blood and feather samples, but only 40.9% in blood and 0% in feathers with P2/P8 (Fitriana *et al.,* 2020; Wirastika *et al.,* 2015). None of those primers top their peers in the sample type they were used. The best primers 2550F/2718R and CHD1F/CHD1R only earned moderate PPI, meaning they are not as universal as the primer with highest PPI. Bali Starling conservation programs should prioritize primers with the highest PPI first before trying taxa specific primer. It will also benefit from technical refinements such as overnight incubation or direct PCR to improve feather DNA yield (Argarini *et al.,* 2020; El Islami *et al.,* 2021; Yuda and Wajjwalku, 2022).

Early and accurate sexing is not a technical luxury but a conservation necessity. Mis-sexed pairs waste breeding seasons and compromise genetic management. By integrating robust primers into Bali Starling workflows, managers can minimize pairing errors, safeguard genetic diversity, and accelerate recovery efforts. As a flagship species, the Bali Starling's adoption of modern molecular sexing sets a precedent for other monomorphic birds in Indonesia, bridging the gap between research and practice.

Limitations

This review is limited by uneven representation across taxa and sample types, with raptors and cockatoos dominating while many families were represented by only one or two studies, restricting generalizability beyond well-sampled groups. Blood and feathers, which accounted for 80% of records, provide the most reliable basis for comparison, whereas eggshell membranes, swabs, and casques were excluded from discussion due to insufficient family and individual numbers that make amplification trends unstable. Primer Performance Index (PPI), although valuable for normalizing success rates across taxa and sample sizes, has inherent constraints. It can inflate reliability when primers are tested broadly but only moderately successful, or undervalue primers with strong raw success in limited taxa. Nuanced methodological differences across studies such as DNA extraction protocols further introduce variability that aggregated metrics cannot capture.

Conclusions

This review shows that molecular sexing using CHD-based PCR is a reliable tool for sexually monomorphic birds in Indonesia, but its success depends on primer selection, sample type, and taxonomic coverage. Blood and feathers remain the most effective samples, while other materials require further validation. Primer performance varied widely across families, indicating that raw success rates alone are insufficient for broad comparison. The Primer Performance Index (PPI) introduced here integrates success rate with taxonomic breadth, identifying CHD1LF/CHD1LR and NP + P2 + MP as the most robust and generalizable primers. Applying these primers can improve breeding management of the critically endangered Bali Starling (*Leucopsar rothschildi*) by reducing sexing errors and optimizing pair formation, supporting stronger integration of molecular sexing into conservation practice.

Acknowledgements. The authors sincerely thank all researchers whose published work contributed to this scoping review. We also appreciate the constructive input and informal discussions that helped refine the ideas presented in this manuscript. This study was conducted without external funding.

References

Adelia, E. I., Srihanto, E. A., Rustiati, E. L., Master, J., Firwantoni, F., Anggy, F. P., & Neli Pratiwi, D. (2025). Sex determination in racing pigeons (*Columba livia*) molecularly using young feather samples. *Jurnal Agrosci, 2*(6), 332–340. https://doi.org/10.62885/agrosci.v2i6.807

Akrom, A. M., Indarjulianto, S., Yanuartono, Y., Susmiati, T., Nururrozi, A., Raharjo, S., Permana, R.G.S., & Sitompul, Y. Y. (2020a). Buccal swabs as sexing material of young nestlings canary bird (*Serinus canaria*) and pigeon (*Columba livia). J. Sain Vet.*, *38*(1), 31-36. https://doi.org/10.22146/jsv.49032

Akrom, A. M., julianto Soedarmanto, I., Nururrozi, A., & Raharjo, S. (2020b). Genotypic sexing in canary (*Serinus canaria*) bird based on chromodomain helicase DNA-binding 1 *(CHD1)* gene. *J. Bioteknol. Biosains Indones.*, *7*(1), 1-8.

Annisa, F., Fikriyanti, M., & Withaningsih, S. (2021). DNA sexing for gender determination of changeable hawk-eagle *(Nisaetus cirrhatus, Gmelin, 1788). E3S Web of Conferences*, *249*, 03012. https://doi.org/10.1051/e3sconf/202124903012

Argarini, A. D., Ari Nugroho, H., Purwaningrum, M., & Haryanto, A. (2020). Molecular bird sexing on fischeri lovebird (*Agapornis fischeri*) by using polymerase chain reaction. *BIO Web Conf.*, *20*, 04003. https://doi.org/10.1051/bioconf/20202004003

Azzahra, L. S., Srihanto, E. A., Rustiati, E. L., Master, J., Firwantoni, F., Febriyani, D. A., & Pratiwi, D. N. (2025). Sex determination on racing pigeons *(Columba livia)* molecularly using blood samples. *J. Agrosci, 2*(6), 323–330. https://doi.org/10.62885/agrosci.v2i6.785

BirdLife International. (2021). *Leucopsar rothschildi*. The IUCN Red List of Threatened Species 2021: e.T22710912A183006359. https://dx.doi.org/10.2305/IUCN.UK.2021-3.RLTS.T22710912A183006359.en. Accessed on 18 May 2026.

Dharmayanthi, A. B., Muchsinin, A., Pulungan, A., & Zein, M. S. A. (2021). Genetic diversity and sex identification of pelicans in Ragunan Zoo, Jakarta. *J. Biol. Indones., 17*(2), 105–114. https://doi.org/10.47349/jbi/17022021/105

El Islami, S. I., Purwaningrum, M., & Haryanto, A. (2021). Molecular sex determination of masked lovebird (*Agapornis personata*) by polymerase chain reaction method. *Proc. KOBI 2nd Int. Confer.*, *1*(1), 48-53.

Fatona, D. R., Permana, R. G. S., Akrom, A. M., Yanuartono, Y., & Indarjulianto, S. (2025). Sexing of young fischer's lovebird (*Agapornis fischeri*) based on chromodomain helicase DNA-binding 1 *(CHD1)* gene and body weight development. *J. Sain Vet., 43*(1), 162. https://doi.org/10.22146/jsv.96593

Fitriana, F., Resita, R., Disastra, Y., Alfatik, G. H., Artdita, C. A., Haryanto, A., & Aziz, F. (2022). Comparison of five types of polymerase chain reaction primers for sex identification in columbidae family birds. *J. Sain Vet., 40*(2), 205-220. https://doi.org/10.22146/jsv.68787

Fitriana, F., Al Qodry, M. F. U. Z., De Lucas, J. C. G., Setyorini, D. R., & Aziz, F. (2023a). Appropriate primer selection improves molecular bird sexing accuracy. *Buletin Peternakan, 47*(4), 215–219. https://doi.org/10.21059/buletinpeternak.v47i4.83320

Fitriana, F., Resita, R., Disastra, Y., Setyorini, D. R., Haryanto, A., & Aziz, F. (2023b). Evaluation of primers targeting chromo helicase DNA-binding gene (CHD) for molecular sexing identification in four bird families. *Livest. Anim. Res., 21*(1), 14-20. https://doi.org/10.20961/lar.v21i1.66998

Fitriana, F., Setyorini, D. R., Artdita, C. A., Ummami, R., Haryanto, A., & Aziz, F. (2023c). Comparison of four types molecular sexing primers in different bird families. *J. Trop. Anim. Vet. Sci., 13*(1), 52-58. https://doi.org/10.46549/jipvet.v13i1.359

Fitriana, Y. S., Irham, M., & Sutrisno, H. (2020). A molecular genetic approach for sex determination on helmeted hornbill (*Rhinoplax vigil)* casque: a forensic casework. *BIO Web Conf., 19*, 00020. https://doi.org/10.1051/bioconf/20201900020

Fuadah, K., Rahmawati, A., Fitriana, F., & Aziz, F. (2025). Evaluation of molecular primers for sexing the magpie robin and green cucak via CHD1 gene amplification. *J. Trop. Anim. Vet. Sci., 15*(2), 64–71. https://doi.org/10.46549/jipvet.v15i2.557

Hall, C. A., Conroy, G. C., & Potvin, D. A. (2025). Ex-situ avian sex skews: determinants and implications for conservation. *PeerJ, 13, e19312.* https://doi.org/10.7717/peerj.19312

Hidayat, R. F. K., Savitri, D., Putri, I., Nugrahani, W. P., & Haryanto, A. (2021). Molecular bird sexing of tanimbar cockatoos (Cacatua goffiniana) by using polymerase chain reaction method. *J. Trop. Biodivers. Biotechnol., 6*(2), 59997. https://doi.org/10.22146/jtbb.59997

Marc, S., Boldura, O. M., Paul, C., Tripon, M. R., Otavă, G., & Savici, J. (2025). Cross-species validation of pigeon-specific CHD1 primers for molecular sexing in pet birds. *Int. J. Mol. Sci., 26*(22), 11142*. https://doi.org/10.3390/ijms262211142*

Nugraheni, P., Purwaningrum, M., Widayanti, R., & Haryanto, A. (2019). Sex determination of peach-faced lovebird *(Agapornis roseicollis)* using polymerase chain reaction (PCR) techniques. *IOP Conf. Ser.: Earth Environ. Sci., 355*(1), 012111. https://doi.org/10.1088/1755-1315/355/1/012111

Nugroho, H. A., & Zein, M. S. A. (2015). Evaluation of sexing methods on black caped lory (*Lorius lory*, Linnaeus 1758). *Zoo Indones.*, *24(2)*, 83-93.

Ó Marcaigh, F., Kelly, D. J., Analuddin, K., Karya, A., Lawless, N., & Marples, N. M. (2021). Cryptic sexual dimorphism reveals differing selection pressures on continental islands. *Biotropica, 53*(1), 121–129. https://doi.org/10.1111/btp.12852

Oktanella, Y. (2017). Sex identification based on molecular marker of CHDZ and CHDW gene in some species of Indonesian birds. *Adv. Anim. Vet. Sci., 7*(10), 844–847. *http://dx.doi.org/10.17582/journal.aavs/2019/7.10.844.847*

Pambuko, G., Vanessa, R., & Prastowo, S. (2023, July). Amplification of *CHD-1* gene fragment in Z and W sex chromosomes of cemani chicken using a different set of PCR primers. In *IOP Conf. Ser.: Earth Environ. Sci.*, *1208*(1), 012058. https://doi.org/10.1088/1755-1315/1208/1/012058

Pamulang, Y. V., & Haryanto, A. (2021). Molecular bird sexing on kutilang (*Pycnonotus* sp.) based on amplification of CHD-Z and CHD-W genes by using polymerase chain reaction method. *Biodiversitas*, *22*(1), 449-452. https://doi.org/10.13057/biodiv/d220155

Petrou, E. L., Scott, L. C., McKeeman, C. M., & Ramey, A. M. (2024). Molecular sexing of birds using quantitative PCR (qPCR) of sex-linked genes and logistic regression models. *Mol. Ecol. Resour., 24*(4), e13946*.* https://doi.org/10.1111/1755-0998.13946

Pinayungan, P. P. W. T., Lovela, A. R., Aksono, E. B., Ismudiono, Setiawan, B., Rachmawati, K., & Rahmahani, J. (2024). Sex identification of the sun conure *(Aratinga solstitialis*) using calamus based on polymerase chain reaction. *Media Kedokt. Hewan, 35*(1), 21–27. https://doi.org/10.20473/mkh.v35i1.2024.21-27

Purwaningrum, M., Nugroho, H. A., Asvan, M., Karyanti, K., Alviyanto, B., Kusuma, R., & Haryanto, A. (2019). Molecular techniques for sex identification of captive birds. *Vet. World*, *12*(9), 1506–1513. https://doi.org/10.14202/vetworld.2019.1506-1513

Ratri, I. N. D., Rahmawati, I. P., Nugrahani, W. P., & Haryanto, A. (2022). Molecular bird sexing of small yellow-crested cockatoo (*Cacatua sulphurea*, Gmelin 1788) using polymerase chain reaction method. *J. Trop. Biodivers. Biotechnol. 7*, 1-12. https://doi.org/10.22146/jtbb.76463

Samad, A., Solihin, D. D., Sumantri, C., & Purwantara, B. (2023). Phylogeography of the maleo senkawor (*Macrocephalon maleo* Sal. Muller 1846) based on cytochrome B gene in Sulawesi and their sex determination using molecular sexing. *Indones. J. Anim. Vet. Sci., 28*(1). https://doi.org/10.14334/jitv.v28i1.3107

Savitri, D., Putri, I., Nugrahani, W. P., Purwaningrum, M., & Haryanto, A. (2021). Molecular bird sexing of sulphur-crested cockatoo *(Cacatua galerita)* by polymerase chain reaction method. *Indones. J. Biotechnol., 26*(1), 1–6. https://doi.org/10.22146/ijbiotech.54611

Sitohang, L. R. (2017). Molecular sexing of Javan hawk-eagle and changeable hawk-eagle confiscated by BKSDA in Yogyakarta. Thesis. Atma Jaya University, Yogyakarta.

Sutomo, Yuni, L. P. E. K., Iryadi, R., & Etten, E. V. (2023). Habitat suitability modeling for Jalak Bali (*Leucopsar rothschildi*) in East Java, Bali, and Lombok: A potential sites for its ex-situ conservation. *AIP Conf. Proc., 6*(1), 050005. https://doi.org/10.1063/5.0118658

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D.,... & Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann. Intern. Med.*, *169*(7), 467-473. https://doi.org/10.7326/M18-0850

Turcu, M. C., Paştiu, A. I., Bel, L. V., & Pusta, D. L. (2023). Minimally invasive sampling methods for molecular sexing of wild and companion birds. *Animals, 13*(21), 3417. https://doi.org/10.3390/ani13213417

Vera, F., Wajjwalku, W., Yuda, P., & Daryono, B. S. (2021). A new primer set in CHD1 gene for bird sex identification. *Biodiversitas, 22*(11), 4977-4982 https://doi.org/10.13057/biodiv/d221133

Wirastika, P. I. P., Yuda, I. P., & Zahida, F. (2015). Sex determination of Bali starling (*Leucopsar rothschildi*) using molecular sexing. *KnE Life Sci., 2*(1), 114. http://dx.doi.org/10.18502/kls.v2i1.128

Withaningsih, S., Ilham, M. F., & Rosdianto, A. M. (2024). Morphometric and DNA sexing accurately in male Javan hawk-eagle (*Nisaetus bartelsi*) determination at Kamojang Eagle Conservation Center, West Java, Indonesia. *Biodiversitas, 25*(3). https://doi.org/10.13057/biodiv/d250332

Yuda, P., & Saputra, A. W. (2021). Eggshell membrane for DNA sexing of the endangered maleo *(Macrocephalon maleo)*. *F1000Research, 9*, 599. https://doi.org/10.12688/f1000research.23712.4

Yuda, P., & Wajjwalku, W. (2022). Using feathers for molecular sexing of straw-headed bulbul (*Pycnonotus zeylanicus*) offsprings. *J. Trop. Biodivers. Biotechnol., 7*(1), 67129. https://doi.org/10.22146/jtbb.67129

Yuda, P., Kinanti, R. A., & Wijaya, A. (2020). Use of swab for DNA sampling from confiscated raptors for molecular sexing. *IOP Conf. Ser.: Earth and Environ. Sci., 590*(1), 012011. https://doi.org/10.1088/1755-1315/590/1/012011

Zein, M. S. A., Haryoko, T., Sulistya Fitriana, Y., Sulistyadi, E., & Prawiradilaga, D. M. (2017). Molecular DNA and phenotypic DNA study application in cockatoo release program. *J. Biol. Indones., 13*(1), 157–169. https://doi.org/10.47349/jbi/13012017/157

STUDIA UNIVERSITATIS BABEŞ-BOLYAI BIOLOGIA, 71, 1, 2026 (p. 39-57)
doi:10.24193/subbbiol.2026.1.03

Doxorubicin toxicity: Histological and ultrastructural tissue damage

Emilia Anca[1], Corina Roşioru[2], Vlad-Alexandru Toma[2,3], Lucian Barbu-Tudoran[2,3,4,5], Septimiu Tripon[4,5], Anca-Daniela Stoica[2,3], and Camelia Dobre[2,3]✉

[1]Doctoral School of Integrative Biology, Faculty of Biology and Geology, Babeş-Bolyai University, Cluj-Napoca, Romania; [2]Faculty of Biology and Geology and Center of Systems Biology, Biodiversity and Bioresources, Babeş-Bolyai University, Cluj-Napoca, Romania; [3]Department of Molecular Biology and Biotechnology, Babeş-Bolyai University, Cluj-Napoca, Romania; [4]Electron Microscopy Laboratory "Prof. C. Crăciun", Faculty of Biology and Geology, Babeş-Bolyai University, Cluj-Napoca, Romania; [5]Electron Microscopy Integrated Laboratory, National Institute for Research and Development of Isotopic and Molecular Technologies, Cluj-Napoca, Romania
✉***Corresponding author, E-mail: camelia.lang@ubbcluj.ro***

Article history: Received 3 April 2025; Revised 1 April 2026; Accepted 2 June 2026; Available online 25 June 2026

Abstract. Doxorubicin (DOX) is an effective chemotherapeutic with known toxic effects over the liver, kidneys and most importantly the heart. DOX effects, such as oxidative stress induction, topoisomerase II poisoning and activation of pro-inflammatory pathways are a double-edged sword. While enhancing the antitumoral activity of this agent, these mechanisms also contribute to tissue damage which limits its use. To build upon previous research that evaluated circulating biomarkers of DOX-induced cardiac toxicity, this study offers additional perspectives on structural effects on tissues along with a short literature review of DOX use.

Keywords: doxorubicin, pharmacological toxicity, histopathology, ultrastructural analysis

Introduction

Anthracyclines are a class of drugs firstly classified as antibiotics at the moment of their discovery in 1939, when daunorubicin was isolated from *Streptomyces peucetius*. Their antitumoral effects were only found over a decade later and represented a groundbreaking achievement for the field of Oncology. This event led to the discovery of several other anthracyclines such as idarubicin, epirubicin and most importantly, doxorubicin (DOX) (Shandilya *et al*., 2020; Xie *et al*., 2024).

General information and uses. DOX, also known as Adriamycin, was first extracted from *Streptomyces peucetius var. caesius* and has been an important drug in the fight against cancer ever since its entry in clinical trials in 1969 (Cassinelli, 2016). It received Food and Drug Administration (FDA) approval in 1974 (Jones and Dass, 2022) and it is included in the World Health Organization (WHO) List of Essential Medicines (WHO, 2023). DOX is used in the treatment of a wide range of cancers, including solid tumors of the breast, bladder, soft tissue and tumors of the hematopoietic and lymphoid tissues, in both pediatric and adult patients (Rawat *et al*., 2021). It is used in veterinary medicine as a cancer treatment as well (Liu *et al*., 2024).

While many DOX formulations have been developed to improve drug efficiency such as liposomes and polymeric-based platforms, few reached the clinical trial stage (Lee *et al*., 2023). In current clinical practice, DOX is used as a hydrochloride salt ("free form") or a liposomal formulation (Li *et al*., 2022).

Liposomal DOX, branded as DoxilÒ, LipodoxÒ or CaelyxÒ, is a nanoformulation consisting of a conjugation of the drug with polyethylene glycol (PEG) which offers a longer circulation time, higher uptake in tumor sites and an enhanced stability of DOX. A non-PEGylated DOX, MyocetÒ is a liposomal formulation containing phosphatidylcholine and cholesterol which also increases its circulation time (Leung *et al*., 2019; van den Boogaard *et al*., 2022).

Other nanoformulations, based on hydrogels, alginate particles and nano-emulsions are constantly being developed and tested. Promising results were obtained by encapsulating DOX in "ferritin cages", which besides the advantages of liposomal formulations, are able to target cancer cells more efficiently (Lee *et al*., 2023; Mattioli *et al*., 2023).

Mechanism of action. The antitumoral effect of DOX is exerted through multiple mechanisms, most of which are not well understood yet. The most potent mechanisms involve DNA damage through intercalation and topoisomerase 2 (TOP2) poisoning, and the overproduction of reactive oxygen species (ROS) which increase oxidative stress and trigger cell death. Additionally, DOX has other less impactful mechanisms that contribute to this effect: senescence and cell

death by iron and calcium homeostasis disruption, and alteration of biological membranes.

Topoisomerase II poisoning and DNA intercalation. TOP2 is an essential enzyme capable of reducing torsional stress created during transcription and replication. These processes create supercoils and tangles in the DNA, and TOP2 fixes them by creating temporary double-strand breaks to allow uncoiling (Yang *et al*., 2014).

DOX intercalates into DNA and blocks the activity of TOP2 right after the double break step and prevents ligation, which activates the DNA damage response. Once activated, the cell cycle is arrested and cell death signaling pathways are initiated. This mechanism is also responsible for some of its side effects (Kciuk *et al*., 2023; van der Zanden *et al*., 2021).

TOP2a is the major isoform of the enzyme and the main target of DOX because of its prevalence in fast-dividing cells, whereas the b isoform is more present in non-dividing or terminally differentiated cells, such as cardiomyocytes where this mechanism contributes to CTOX (Yang *et al*., 2014).

DOX-DNA adducts formation and chromatin damage. This mechanism has a strong relationship to TOP2 poisoning but is treated in literature as a separate mechanism. The binding of DOX into the DNA helix activates damage responses itself and is facilitated by formaldehyde that comes from the interaction between free radicals and lipids. Although adduct formation does not have a meaningful impact on cancer cells, the addition of formaldehyde-forming compounds to the chemotherapeutic regimen was proven beneficial (Yang *et al*., 2014).

Another effect of DOX-DNA complexes is chromatin damage through histone eviction which leads to mostly epigenomic changes, but the transcriptional landscape modifications are notable as well. An example of this mechanism consists in the expression of oxidative stress- and cell death-related genes (Jones and Dass, 2022; van der Zanden *et al*., 2021).

ROS-induced damage. Oxidative stress represents an imbalance between ROS production and the activity of the antioxidant defense system - comprised of enzymes and antioxidants (Shi *et al*., 2023).

ROS are fundamental for cell signaling processes and immune system function, if present at physiological levels and given that the cell's antioxidant system function properly (Forman *et al*., 2010). DOX metabolization leads to the overproduction of ROS, effect that can be imagined as a double-edged sword. While ROS and increased oxidative stress are crucial in the antitumoral effect of DOX, these also lead to cellular injury in healthy tissues and notorious side effects that limit its use.

DOX has a high affinity for cardiolipin, a lipid found in the inner mitochondrial membrane. The interaction between these molecules inhibits the activity of the

electron transport chain (ETC) and accumulates DOX in this cell compartment. Oxidoreductases found here produce superoxide radicals from the quinone moiety of DOX, which are converted by superoxide dismutase (SOD) into hydrogen peroxide (H_2O_2), to reduce its damaging power. In the presence of Fe^{2+} (which accumulates in the mitochondria due to DOX), H_2O_2 is converted into the more potent hydroxyl radical. Because this cannot be detoxified by the enzymatic antioxidants, its presence leads to nucleic acid damage and lipid peroxidation (Shi *et al*., 2023). The peroxidation of lipid membranes increases the concentration of aldehydic compounds, which in turn produce more ROS and activates cell death pathways (Christidi and Brunham, 2021).

DOX-induced cell death. Autophagy is a response triggered by cell stress and consists of cell degradation followed by the recycling of useful components in the process. In DOX treatment, this is stimulated by mitochondrial damage and oxidative stress and is considered an antitumoral mechanism if cellular DOX concentration is under a toxic threshold. High or chronic exposure is associated with suppression of this process which is known to promote tumor development and drug resistance (Kciuk *et al*., 2023).

Necroptosis, another form of cell death is mediated by tumor necrosis factor alpha (TNFa) signaling, which activates the formation of the necroptosome complex. This leads to cell membrane rupture and the release of cellular components that trigger immune responses (Christidi and Brunham, 2021).

Ferroptosis is an iron-dependent form of cell death that is strictly tied to ROS accumulation. DOX, and its main metabolite, doxorubicinol stop iron metabolism by making iron metabolism proteins inactive, thus encouraging iron pooling (Rawat *et al*., 2021). The mitochondrial iron accumulated by DOX catalyzes the production of ROS as described above. This type of cell death is also strictly related to lipid peroxidation and is responsible for part of the cardiac damage produced by DOX (Christidi and Brunham, 2021).

Pyroptosis or immunogenic cell death is an inflammation-promoting type of cell death induced by DOX binding to gasdermin D and E that further activates caspases and cell death (Vitale *et al*., 2024). Simultaneously, this effect can trigger T-cell immune responses which can attack tumor cells. Although undesired for the implication in CTOX, this consequence of DOX could be used to enhance the efficiency of immunotherapies when administered in the same treatment regimen (Gabizon *et al*., 2025).

Cell death initiated by calcium homeostasis dysregulation is produced by doxorubicinol, which similarly to ferroptosis, is able to increase the intracellular Ca^{2+} levels by inhibiting the sodium-calcium exchanger protein. In muscle cells, calcium overload activates calpains which activates caspases and increase ROS production, thus leading to cell death (Rawat *et al*., 2021).

Cardiotoxicity. In the case of DOX administration, CTOX is the most daunting side effect, especially in childhood cancer survivors. Pediatric patients can often tolerate more aggressive treatment regimens, most likely due to fewer comorbidities, better organ function and different drug metabolism than adults. Yet, this population is more affected by tardive side effects of cancer treatments that can alter their quality of life, and physicians should be more aware of warning signs of side effects than in adults (Helms *et al*., 2023; van den Boogaard *et al*., 2022).

Medical practitioners became aware of the cardiac damage produced by anthracyclines shortly after daunorubicin was introduced in clinical use. In the two "Letters to the Editor" of The Lancet journal in 1969, authors described their cases and outcomes, and advised their colleagues on methods to prevent CTOX, mainly by reserving daunorubicin use only for young patients affected by leukemia and avoiding combined therapies with drugs that increase the cardiotoxic effects (Bonadonna and Monfardini, 1969).

Symptoms of CTOX are often "silent" at the onset of the disease or can go unnoticed because they are non-specific. This is the case for intermittent chest pain, fatigue, shortness of breath, vertigo, heart palpitations or limb swelling because in the absence of cardiovascular diagnostic tests, these can be easily attributed to other conditions (Rawat *et al*., 2021).

A challenging fact to acknowledge is that gender bias heavily impacts CVD diagnosis. This is proven by the fact that mortality rate is higher in women compared to men (Townsend *et al*., 2022). Although several factors such as smoking, obesity, menopause and pregnancy are contributors to risk increase, these should not affect the diagnostic process (Sliwa *et al*., 2021). An increasing pool of evidence shows that CVD symptoms are often disregarded or misinterpreted by medical practitioners as anxiety, depression, or just fatigue in general. These misconceptions, along with the tendency to understate the severity of their symptoms lead to less women receiving referrals to a cardiologist (Al Hamid *et al*., 2024; Berg Gundersen *et al*., 2017; Ezekowitz *et al*., 2020; Sliwa *et al*., 2021).

Despite extensive research, the mechanism of CTOX development is still unclear, especially in the case of late-onset or chronic CTOX, occurring years after treatment completion. For acute forms of the disease, tissue uptake of DOX increases oxidative stress, thus promoting cell damage and apoptotic pathways.

In the case of CTOX occurring long after treatment, the only evidence-supported hypothesis is that DOX treatment promotes an aging phenotype in heart tissue, which cannot be identified by current diagnosis methods (Linders *et al*., 2024).

As described in Chapter II, a method for preventing or curing CTOX is not yet available. The most clinically effective solution seems to stand in a combination of both conventional and traditional medicine, along with a non-invasive

screening of patients to identify subclinical CTOX swiftly and implement treatment as soon as possible.

Materials and Methods

The study adhered to the requirements of the European Directive 2010/63/EU and national legislation 43/2014 regarding the protection of animals used for scientific purposes. The protocol was approved by the Committee on the Ethics of Animal Experiments of Babeş-Bolyai University (14.172/02.11.2021) and was compliant to the ARRIVE guidelines. Adult male Wistar rats (of average weight 250 g at the beginning of the experiment) were housed in the Animal Laboratory Facility (Zoobase) at the Babeş-Bolyai University in Cluj-Napoca. Throughout the experiment, animals received free access to standard diet and water, were gently handled and were kept in standard cages on a 12-hour light-dark cycle.

The treatment protocol has been described previously (Anca *et al*., 2024). Briefly, a cumulative dose of 15 mg/kg DOX hydrochloride was divided into 4 weekly doses and administered intravenously via tail vein. Control group was treated with vehicle, in the same manner. Tissues were harvested post-treatment and at 4 weeks post-treatment for the acute and chronic manifestations of toxicity, respectively.

For ultrastructural analysis, the protocol described by Crăciun & Barbu-Tudoran (2013). Briefly, collected heart and liver tissues were quicky fixed in 2.7% glutaraldehyde, washed and post-fixed in 2% osmium tetraoxide. After acetone dehydration, samples were embedded in Epon 812 resin (Electron Microscopy Science USA). Section trimmings were performed using a Leica UC6 microtome, and then sections were placed on electron microscope grids. After double contrasted using uranyl acetate (UranyLess, Electron Microscopy Science, USA), grids were examined using a Jeol JSM 1010 (Japan) transmission electron microscope. Images were post-edited for contrast using Adobe Photoshop software. Transmission electron microscopy images were obtained at the "Constantin Crăciun" Electron Microscopy Center.

For histopathological analysis, heart, aorta, liver and kidney sections were prepared for hematoxylin-eosin (H&E) staining. Briefly, tissue fragments were isolated and fixed in a 5% neutral formalin solution, for approximately 72 hours. After paraffin embedding, multiple sections were cut at 5 µm thickness and mounted onto glass slides. After dewaxing in xylene and rehydration, sections were stained using the standard H&E procedure. Slides were then investigated using an Optika 510 LD1 microscope with a 5MP charge-coupled device (CCD) camera.

Results

Ultrastructural analysis of heart tissue from Control rats shows normal cell architecture at both time-points, with clearly delineated myofibrils running parallel to each other. Sarcomeres are visibly separated by Z membranes. Mitochondria are electron dense and contain unaltered cristae. The euchromatic nucleus contains a well-defined nucleolus (Fig. 1A-D).

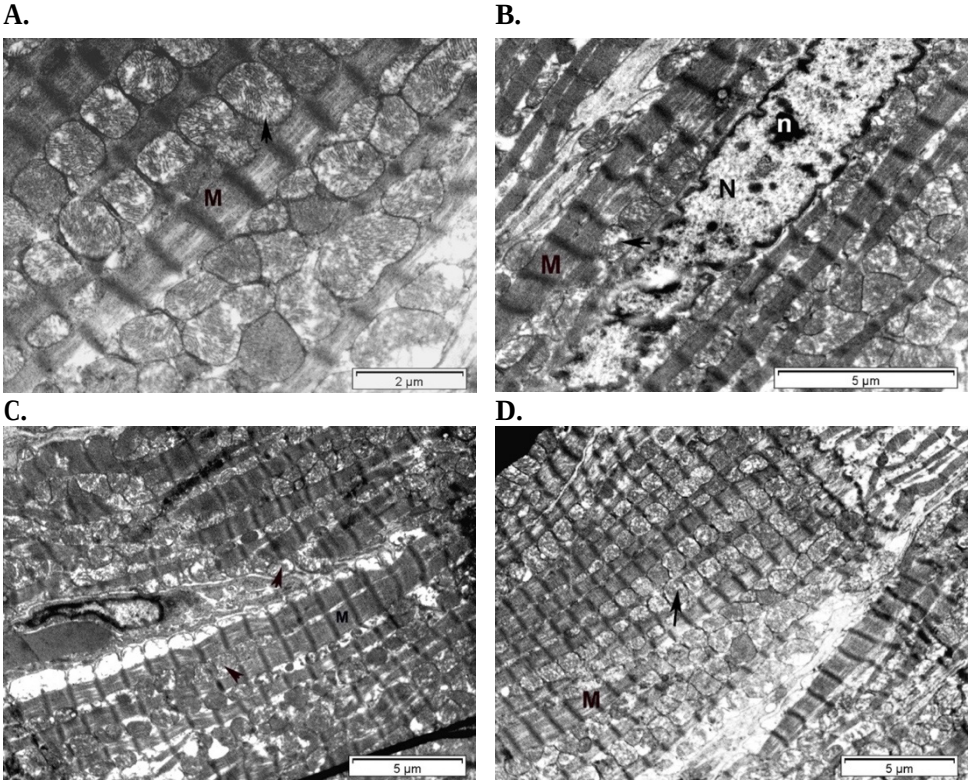


Figure 1. Electron micrographs of Control group myocardial tissue, post-treatment (A, B) and 4 weeks post-treatment (C, D). A. Well-defined myofibrils (M), running parallel to each other, with visible sarcomeres delimited between two neighboring Z membranes; B. between myofibrils, rows of mitochondria are observed (marked by arrows); these mitochondria are normally electron dense, with intact cristae. The nucleus (N) is euchromatic, with well-differentiated nucleolus (n); C. normal cardiomyocytes, with normally arranged myofibrils (M) in parallel rows; D. normal arrangement of mitochondria (marked by arrows), parallel to myofibrils.

DOX group ultrastructural analysis showed extensive damage at both time-points, with disorganized myofibrils, collapsed mitochondrial cristae and numerous lipid and collagen deposits (Fig. 2A-B). Post-treatment, nuclei were predominantly

heterochromatic, with irregular shapes and surrounded by vacuolization processes. Interestingly, at this time-point we mainly observed damaged mitochondria, which appear to fuse at 4 weeks post-treatment (Fig. 2C-D).

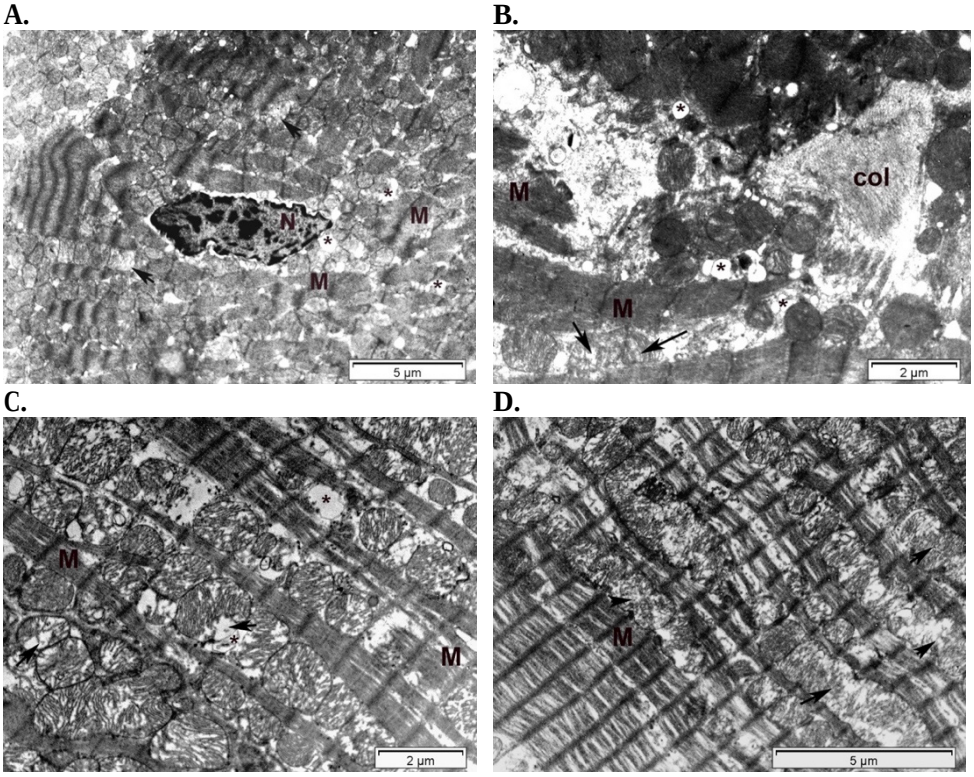


Figure 2. Electron micrographs of DOX group myocardial tissue, post-treatment (A, B) and 4 weeks post-treatment (C, D). A. disorganized myofibrils (M), mitochondria (marked by arrows) with collapsed cristae, heterochromatic nucleus (N) of irregular shape and perinuclear cytoplasmic vacuolisation; B. intracellular lipid droplets (marked by *) and collagen deposits; C. degraded myofibrils (M), with denatured content; disorganized mitochondrial cristae (arrow); lipid droplets in between mitochondria (marked by *); D. some mitochondria show fusion-like behavior, as a response to intense metabolic stress.

All histological samples were collected only after the completion of the 4-week course of treatment. Histological analysis of heart sections from the Control group show intact tissue structure (Fig. 3A-B). In DOX group, disorganized muscle bundles along with leukocyte infiltrates and visible microhemorrhages (MHs) are seen (Fig. 3C-D). Additionally, post-treatment cardiomyocytes show excentric nuclei position and rarefied cytoplasm (Fig. 3E).

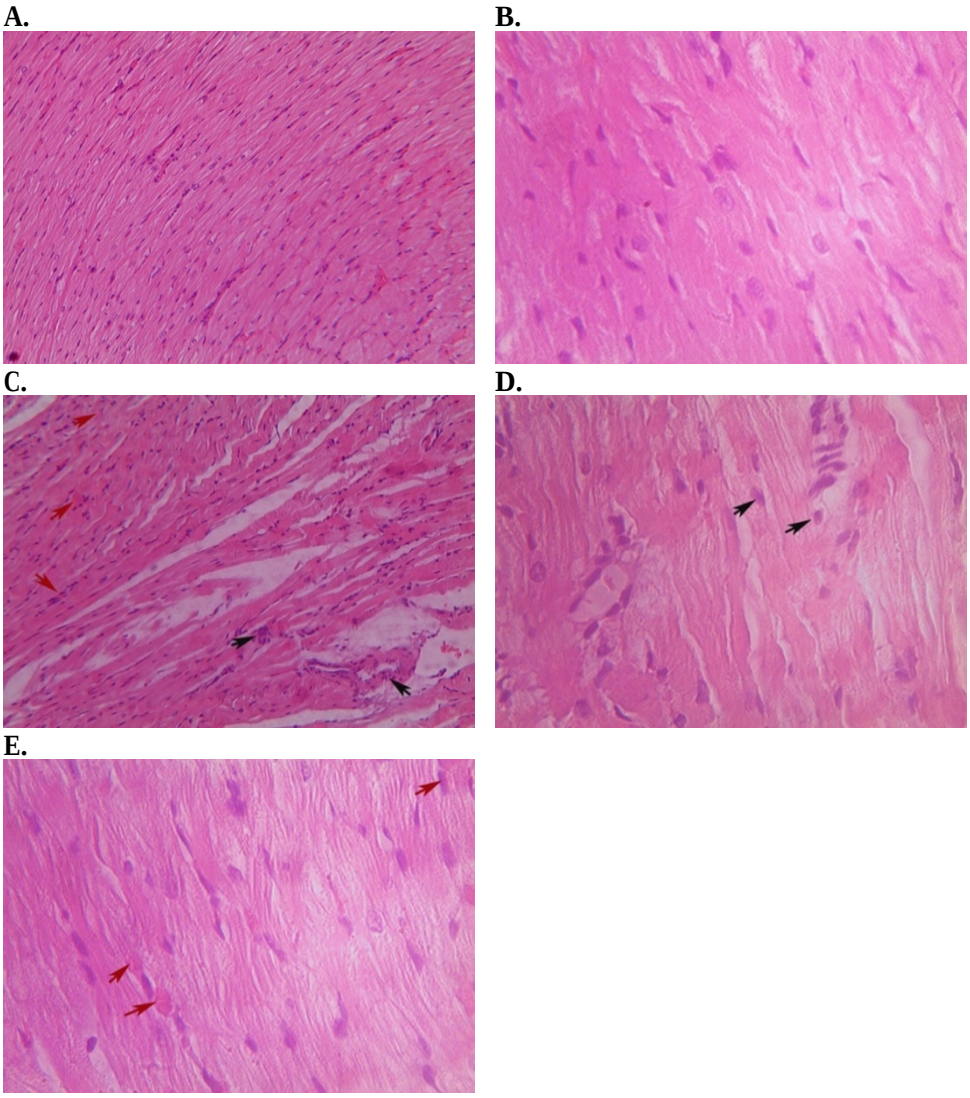


Figure 3. Heart sections of Control and DOX groups, H&E staining. A and B. Control post treatment. Both images are showing normal tissue architecture x10 for A, x40 for B; C. DOX, post-treatment, x10; D and E. DOX, post-treatment, x40. Black arrows indicate leukocyte infiltrates and red arrows indicate MH.

Control group aorta samples show normal cell and tissue structure. In tunica media, rows of smooth muscle cells are intercalated with parallel elastic fibers (Fig. 4A). DOX treatment caused a relative thinning of the tunica media and thickening of the tunica adventitia (Fig. 4B).

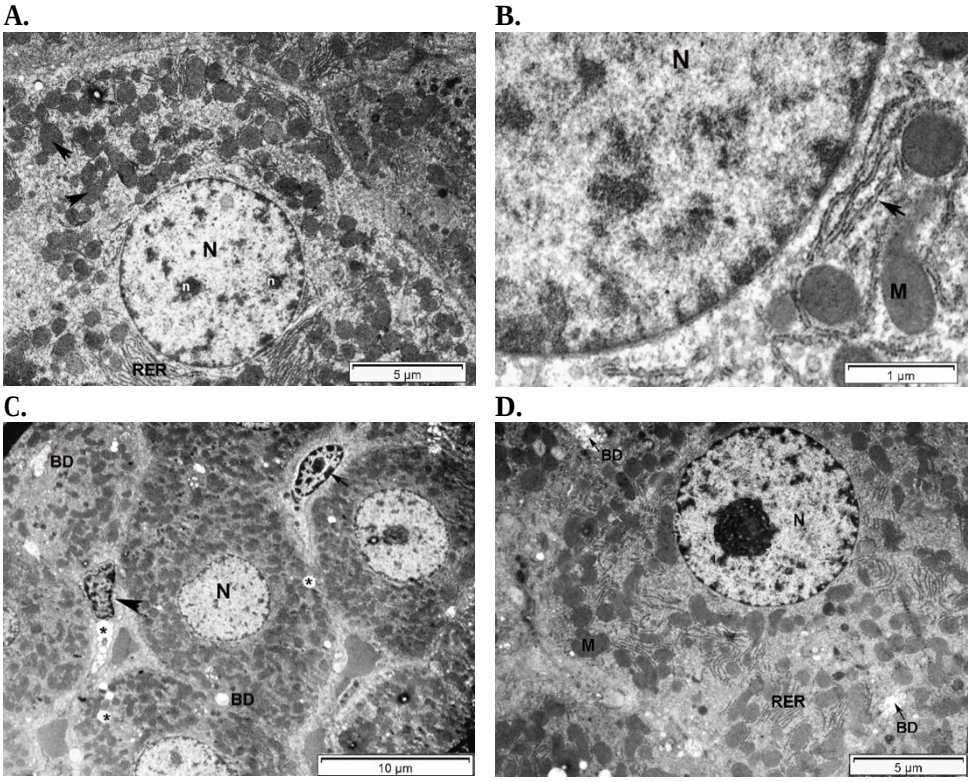


Figure 4. Electron micrographs of Control group liver tissue, post-treatment (A, B) and 4 weeks post-treatment (C, D). A. hepatocyte with euchromatic, spherical nucleus (N) and 2 nucleoli (n), and parallel RER. Mitochondria (arrows) are normal electron dense; **B.** mitochondrial complexes surrounded by RER (arrow); **C.** parenchymal cells with central, well-defined nucleus (N), hepatocyte microvilli present in the BD and in the Disse spaces, lipid droplets (marked by *) in hepatocytes and Ito cells (thick arrow), and a Kupffer cell (thin arrow); **D.** hepatocyte with euchromatic nucleus and large nucleolus, RER and mitochondrial complexes, and BD. RER – rough endoplasmic reticulum; BD –bile duct.

Control group TEM images for both time-points show normal cellular architecture. Post-treatment, hepatocytes have euchromatic, spherical nuclei surrounded by numerous cisternae of the rough endoplasmic reticulum (RER) and mitochondria (Fig. 4A-B). Additionally, at 4 weeks post-treatment, few lipid droplets can be observed in hepatic stellate cells and hepatocytes (Fig. 4B-C). At both time-points, DOX led to SER and RER proliferation (Fig. 5AD), and hepatocyte apoptosis (Fig. 5B-C), with more cellular damage observed at 4-week post-treatment (Fig. 5E).

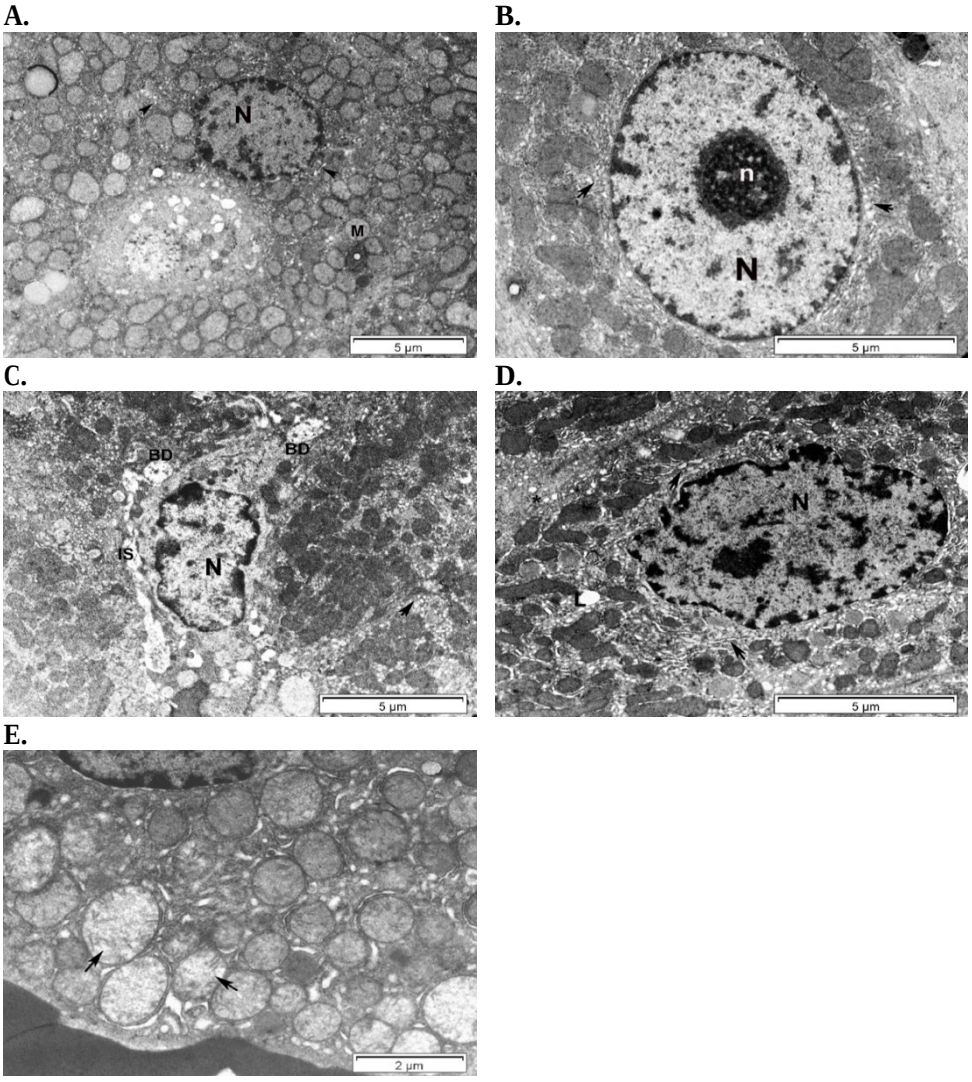


Figure 5. Electron micrographs of DOX group liver tissue, post-treatment (A, B) and 4 weeks post-treatment (C, D, E). A. dilated membrane vesicles of SER (arrows); cellular debris as remnants of a probably apoptotic cell; **B.** cell with intact nucleus (N) and nucleolus (n); massive proliferation of endoplasmic reticulum; **C.** hepatocyte with irregular nuclear (N) shape and enlarged intercellular space (IS), and bile ducts with altered hepatocyte microvilli (BD), proliferation of SER (arrow); **D.** hepatocyte with heterochromatic irregular nucleus (N), proliferation of SER and RER (arrow and *, respectively); **E.** faulty mitochondria, with damaged cristae (arrows) between dilated RER profiles. SER – smooth endoplasmic reticulum; RER – rough endoplasmic reticulum.

To further evaluate the extent of structural disruption caused by DOX, histological sections were examined. Micrographs of livers from the Control group show physiological tissue structure, with normal hepatocytes, arranged in continuous plates around the centrilobular vein (CV) (Fig. 6A-B). DOX treatment led to loss of normal tissue architecture, with enlargement of the CV and dilated sinusoids, along with inclusions between hepatocytes (Fig. 6C). At week 4 post-treatment, we observed the proliferation of biliary ducts, along with enlarged Disse spaces and disorganized cytoplasm in hepatocytes (Fig. 6D-E).

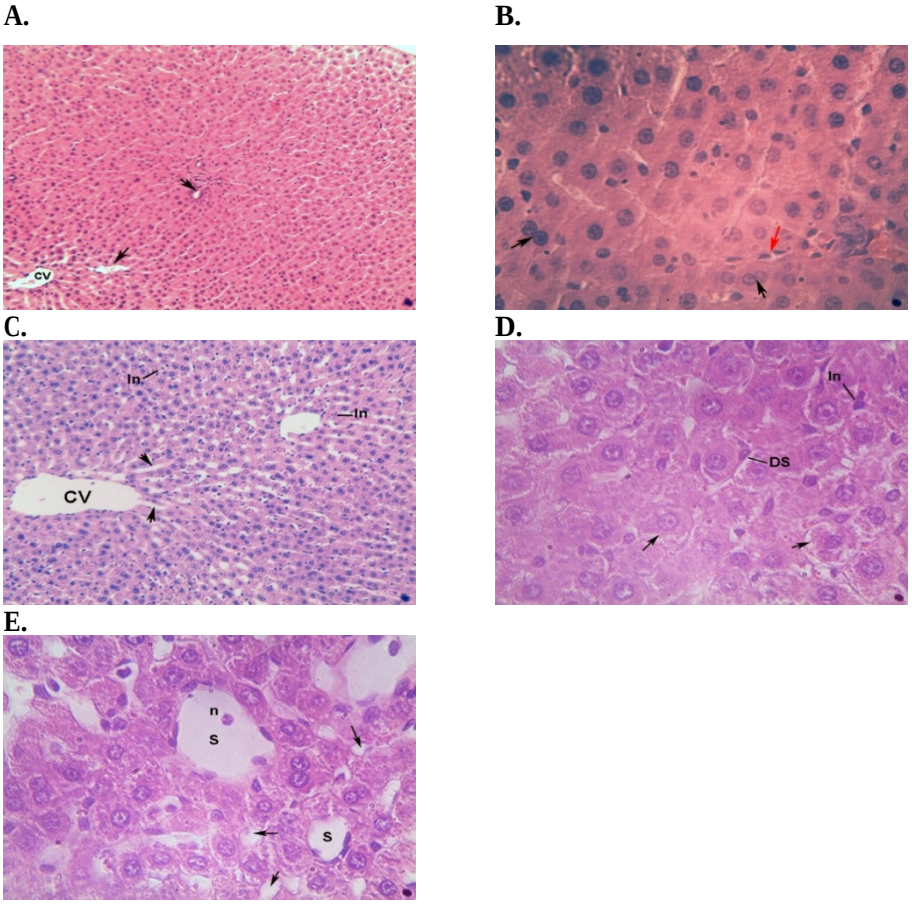


Figure 6. Liver sections of Control and DOX groups, H&E staining. A and B. Control post-treatment images, showing intact tissue structure. Centrilobular vein (CV), sinusoid capillaries (black arrows in A), Kupffer cell (red arrow in B), binucleated cells (black arrows in B); x10 for A, x40 for B; C. DOX, post-treatment image with CV, sinusoid dilatation (arrows) and inclusion bodies (In), x10; D. DOX post-treatment, indicating disorganized cytoplasm marked by arrows, enlarged Disse spaces (DS) and cell inclusions (In), x40; E. DOX post-treatment, S indicates liver sinusoids, proliferated biliary ducts (arrows), a neutrophil in a sinusoid capillary (n), x40.

Kidney photomicrographs for the Control group show normal tissue architecture, with round, well-conturated glomeruli (G) and visible macula densa (MD). Proximal (PT) and distal convoluted tubes (DT) are normal appearing (Fig. 7A-B). DOX group, similar to the other organs analyzed, shows irreversible structural damage. Vascular injury is reflected by multiple MH spots visible at both magnifications used. Some nuclei have pyknotic appearance and are displaced beyond the cellular boundary, which indicates active apoptotic or necrotic processes. The majority of renal corpuscles were severly deformed (Fig. 7C-E).

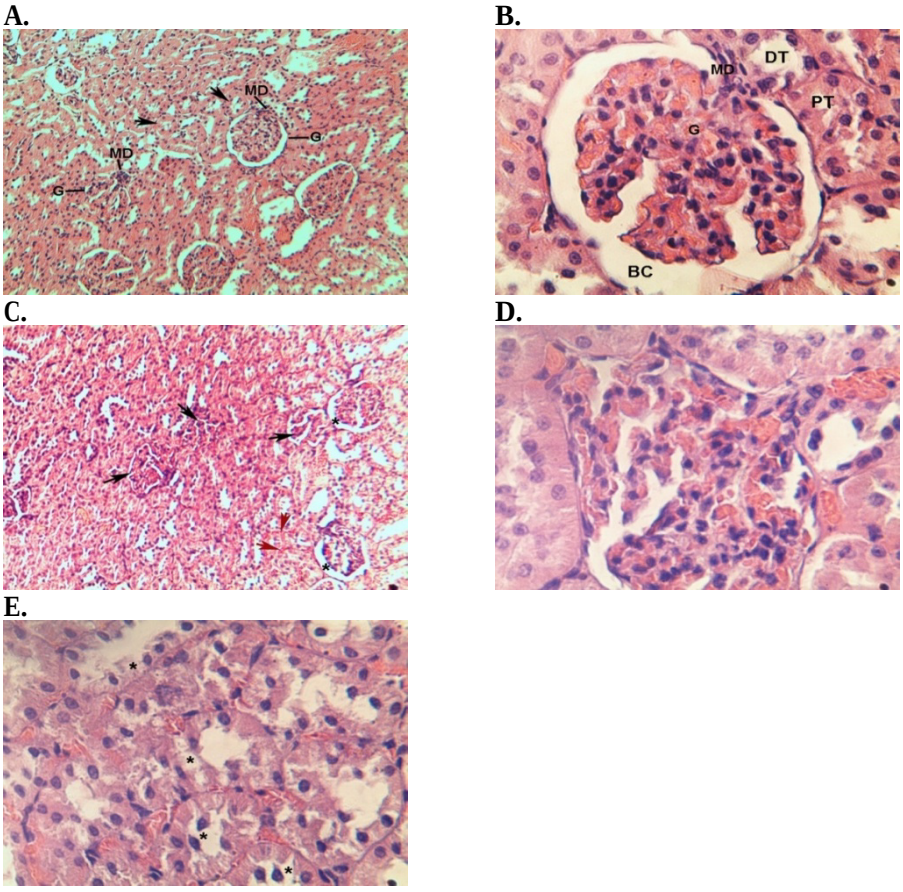


Figure 7. Kidney sections of Control and DOX groups, H&E staining. A. In Control group post-treatment, arrows indicate normal tubular epithelium (arrows) and glomeruli (G); at some glomeruli, the macula densa (MD) can also be seen; x10; B. A round renal corpuscle in the Control group with normal Bowman capsule (BC) and capillary loops of the glomerule (G); the macula densa (MD) is visible at the beginning of distal tube (DT), near a loop of the proximal tube (PT); x40; C. DOX group, post-treatment. The renal corpuscles (G, black arrows) are deeply altered, with missing, or enlarged Bowman

capsule (*); the red arrows indicate MH foci; x10; D. A renal corpuscle in the DOX-treated group. The Bowman capsule is not visible and the capillary loops of the glomerule are disorganized; x40. E. DOX group, post-treatment. Renal tubes with altered cells, some of them with disrupted membrane and nuclei beyond the cell boundaries (*); x40.

Discussion

Cardiomyocytes are mitochondria-rich cells, due to their intense need for energy. Under improper conditions such as extreme oxidative stress, mitochondria undergo inner and outer membrane fusion. As a result, the fused mitochondria share their contents to maintain normal function (Adebayo *et al*., 2021). This quick process is mediated by mitofusins (MFN1 and MFN2) for the outer membrane, whereas the inner mitochondrial membrane is fused by optic atrophy 1 (OPA1). Errors in this process are lethal, as proven by mutant mice experiments (Pernas and Scorrano, 2016).

As DOX is an inhibitor of mitochondrial fusion and fission (Ding *et al*., 2022), this process was not observed at the end of treatment (Fig. 2AB). However, as shown in Fig. 4B, we observed fused mitochondria at 4 weeks post-treatment. This suggests that after DOX exposure ends, although cardiomyocytes show clear signs of severe energy impairment, partially impaired mitochondria are adapting to the altered cell environment and compensating for the damage.

Collagen accumulation in cardiac muscle has been observed in DOX treated mice (He *et al*., 2020b) and is a sign of hypertrophy and fibrosis (Levick *et al*., 2019). One of the ways DOX could intensify collagen production and deposition is by upregulation of Adamts19, which we observed in our transcriptomic analysis of heart tissue after the same DOX protocol (results not published).

Morphological evaluation of heart sections revealed oxidative-stress related tissue injury, which presents with separation of myofibers, inflammatory cell infiltration and microhemorrhages (MH) (Fig. 3 D-E).

The effects DOX exhibits on the liver need as much attention as the heart, because the oxidative stress produced by DOX metabolism greatly impacts the liver (Gedikli *et al*., 2023). DOX heavily impacts the liver by decreasing lipid metabolism which ultimately leads to nonalcoholic steatohepatitis (Timm *et al*., 2022). DOX-induced liver injury is led by ROS overproduction, which lowers the activity of the antioxidant molecules (glutathione peroxidase, superoxide dismutase, catalase and glutathione) (Prasanna *et al*., 2020).

In our previous study, we demonstrated that DOX produces hyperlipidemia which continues to worsen even after treatment completion (Anca *et al*., 2024). Given that liver dysfunction is a widely recognized risk factor for cardiovascular disease (Lee *et al*., 2020), these results consolidate the importance of early protective strategies for patients undergoing chemotherapy regimens that include DOX.

Hepatocyte ballooning, which is characterized by enlarged cells with dilated cytoplasm, is a sign of injury and metabolism alteration as a response to DOX-induced oxidative stress (Bonnet *et al*., 2022; Fatani *et al*., 2022). This was observed in post-treatment liver sections (Fig. 7 D-E), which may also predict a prolonged alteration of the tissue following DOX exposure. Additionally, we observed numerous MH spots caused by sinusoidal endothelial barrier integrity disruption (Fig. 7C-E). This type of endothelial injury could be attributed to DOX-induced mitochondrial dysfunction: excessive ROS production in the mitochondria can trigger a ROS-induced ROS release mechanism, which in turn damages the endothelial wall (He *et al*., 2020a).

The smooth ER is an important player in lipid metabolism, especially in the liver. ER proliferation is a known stress marker associated with cardiovascular and metabolic diseases (Sozen and Ozer, 2017) by increasing fatty acid synthesis (Zhou *et al*., 2022).

Histological analysis of kidney tissue revealed severe and progressive renal damage. As indicated by multiple MH spots, vascular injury seems to be an ongoing process (Fig. 8C). Additionally, the altered appearance of the renal corpuscles adjacent to the MH and displacement of the nuclei beyond the cellular membrane collectively highlight the degree of DOX toxicity in the kidney (Fig. 8BC). As in the other tissues, DOX-kidney injury is likely caused by oxidative stress and inflammation (Hussain *et al*., 2021; Szalay *et al*., 2015).

Conclusion

The present work provides additional information regarding histological and ultrastructural effects of DOX on tissues to build upon previous research on the evaluation of circulating biomarkers. The systemic impact of DOX is manifested through oxidative stress, vascular injury and lipid dysregulation.

In all organs analyzed, DOX led to progressive tissue dysfunction manifested as apoptotic-like cell appearance, myofibrillar disarray and vascular damage and immune cell infiltration. These results, combined with our previous observations, emphasize the need of a multifaceted therapeutic approach to DOX toxicity centered on reduction of oxidative stress to non-target organs.

References

Adebayo, M., Singh, S., Singh, A.P. & Dasgupta, S. (2021). Mitochondrial fusion and fission: the fine-tune balance for cellular homeostasis. *FASEB J. 35(6)*, 321620. https://doi.org/10.1096/fj.202100067R

Al Hamid, A., Beckett, R., Wilson, M., Jalal, Z., Cheema, E., Al-Jumeily OBE, D., Coombs, T., Ralebitso-Senior, K. & Assi, S. (2024). Gender bias in diagnosis, prevention, and treatment of cardiovascular diseases: a systematic review. *Cureus 16(2)*, e54264. https://doi.org/10.7759/cureus.54264

Anca, E., Sabău, F., Vădan, A., Marinescu, M., Licărete, E., Roșioru, C., Stoica, A.D., Dobre, C. & Banciu, M. (2024). Assessment of circulating biomarkers in a rat model of doxorubicin-induced cardiotoxicity. *Stud. Univ. Babes Bolyai Biol. 69(2)*, 7–22. https://doi.org/10.24193/subbbiol.2024.2.01

Berg Gundersen, A.E., Sørlie, T. & Bergvik, S. (2017). Women with coronary heart disease – making sense of their symptoms and their experiences from interacting with their general practitioners. *Health Psychol. Behav. Med. 5*, 29–40. https://doi.org/10.1080/21642850.2016.1263574

Bonadonna, G. Monfardini, S. (1969). Cardiac toxicity of daunorubicin. *Lancet 293(7599)*, 837-838. https://doi.org/10.1016/s0140-6736(69)92093-5

Bonnet, L., Alexandersson, I., Baboota, R.K., Kroon, T., Oscarsson, J., Smith, U. & Boucher, J. (2022). Cellular senescence in hepatocytes contributes to metabolic disturbances in NASH. *Front. Endocrinol. (Lausanne) 13*, 957616. https://doi.org/10.3389/fendo.2022.957616

Cassinelli, G. (2016). The roots of modern oncology: from discovery of new antitumor anthracyclines to their clinical use. *Tumori 2016(3)*, 226-235. https://doi.org/10.5301/tj.5000507

Christidi, E. & Brunham, L.R. (2021). Regulated cell death pathways in doxorubicin-induced cardiotoxicity. *Cell Death. Dis. 12(4)*, 339. https://doi.org/10.1038/s41419-021-03614-x

Craciun, C. & Barbu-Tudoran, L. (2013). Identification of new structural elements within "porosomes" of the exocrine pancreas: a detailed study using high-resolution electron microscopy. *Micron 44*, 137-142. https://doi.org/10.1016/j.micron.2012.05.011

Ding, M., Shi, R., Cheng, S., Li, M., De, D., Liu, C., Gu, X., Li, J., Zhang, S., Jia, M., Fan, R., Pei, J. & Fu, F. (2022). Mfn2-mediated mitochondrial fusion alleviates doxorubicin-induced cardiotoxicity with enhancing its anticancer activity through metabolic switch. *Redox Biol. 52*, 102311. https://doi.org/10.1016/j.redox.2022.102311

Ezekowitz, J.A., Savu, A., Welsh, R.C., McAlister, F.A., Goodman, S.G. & Kaul, P., (2020). Is there a sex gap in surviving an acute coronary syndrome or subsequent development of heart failure? *Circulation 142(23)*, 2231-2239. https://doi.org/10.1161/CIRCULATIONAHA.120.048015

Fatani, A., Baothman, O., Shash, L., Abuaraki, H., Zeyadi, M., Hosawi, S., Altayb, H. & Abo-Golayel, M. (2022). Hepatoprotective effect of date palm fruit extract against doxorubicin intoxication in Wistar rats: in vivo and in silico studies. *Asian Pac. J. Trop. Biomed. 12(8)*, 357-366. https://doi.org/10.4103/2221-1691.350184

Forman, H.J., Maiorino, M. & Ursini, F. (2010). Signaling functions of reactive oxygen species. *Biochemistry. 49(5)*, 835-842. https://doi.org/10.1021/bi9020378

Gabizon, A.A., Gabizon-Peretz, S., Modaresahmadi, S. & La-Beck, N.M. (2025). Thirty years from FDA approval of pegylated liposomal doxorubicin (Doxil/Caelyx): an updated analysis and future perspective. *BMJ Oncol. 4(1)*, e000573. https://doi.org/10.1136/bmjonc-2024-000573

Gedikli, E., Barış, V.Ö., Yersal, N., Dinçsoy, A.B., Müftüoğlu, S.F. & Erdem, A. (2023). Taurine protects doxorubicin-Induced hepatotoxicity via its membrane-stabilizing effect in rats. *Life 13(10)*, 2031. https://doi.org/10.3390/life13102031

He, H., Wang, L., Qiao, Y., Zhou, Q., Li, H., Chen, S., Yin, D., Huang, Q. & He, M., (2020a). Doxorubicin induces endotheliotoxicity and mitochondrial dysfunction via ROS/eNOS/NO pathway. *Front. Pharmacol. 10*, 1531. https://doi.org/10.3389/fphar.2019.01531

He, Q., Wang, F., Ryan, T.D., Chalasani, M. & Redington, A.N. (2020b). Repeated remote ischemic conditioning reduces doxorubicin-induced cardiotoxicity. JACC *CardioOncol. 2(1)*, 41-52. https://doi.org/10.1016/j.jaccao.2020.01.005

Helms, L., Guimera, A.E., Janeway, K.A. & Bailey, K.M. (2023). Innovations in cancer treatment of children. *Pediatrics 152(6)*, e2020301539. https://doi.org/10.1542/peds.2023-061539

Hussain, M.A., Abogresha, N.M., Abdelkader, G., Hassan, R., Abdelaziz, E.Z. & Greish, S.M. (2021). Antioxidant and anti-inflammatory effects of crocin ameliorate doxorubicin-induced nephrotoxicity in rats. *Oxid. Med. Cell Longev. 2021*, 8841726. https://doi.org/10.1155/2021/8841726

Jones, I.C. & Dass, C.R. (2022). Doxorubicin-induced cardiotoxicity: causative factors and possible interventions. *J. Phar. Pharmacol. 74(12)*, 167-1688. https://doi.org/10.1093/jpp/rgac063

Kciuk, M., Gielecińska, A., Mujwar, S., Kołat, D., Kałuzińska-Kołat, Ż., Celik, I. & Kontek, R. (2023). Doxorubicin—an agent with multiple mechanisms of anticancer activity. *Cells 12(4)*, 659. https://doi.org/10.3390/cells12040659

Lee, J., Choi, M.K. & Song, I.S., (2023). Recent advances in doxorubicin formulation to enhance pharmacokinetics and tumor targeting. *Pharmaceuticals 16(6)*, 802. https://doi.org/10.3390/ph16060802

Lee, S.F., Luque-Fernandez, M.A., Chen, Y.H., Catalano, P.J., Chiang, C.L., Wan, E.Y.F., Wong, I.C.K., Chen, M.H. & Ng, A.K. (2020). Doxorubicin and subsequent risk of cardiovascular diseases among survivors of diffuse large B-cell lymphoma in Hong Kong *Blood Adv. 4(20)*, 5107-5117. https://doi.org/10.1182/BLOODADVANCES.2020002737

Leung, A.W.Y., Amador, C., Wang, L.C., Mody, U. V. & Bally, M.B. (2019). What drives innovation: the canadian touch on liposomal therapeutics. *Pharmaceutics 11(3)*, 124 https://doi.org/10.3390/pharmaceutics11030124

Levick, S.P., Soto-Pantoja, D.R., Bi, J., Hundley, W.G., Widiapradja, A., Manteufel, E.J., Bradshaw, T.W. & Meléndez, G.C. (2019). Doxorubicin-induced myocardial fibrosis involves the neurokinin-1 receptor and direct effects on cardiac fibroblasts. *Heart Lung Circ. 28(10)*, 1598-1605 https://doi.org/10.1016/j.hlc.2018.08.003

Li, X.R., Cheng, X.H., Zhang, G.N., Wang, X.X. & Huang, J.M. (2022). Cardiac safety analysis of first-line chemotherapy drug pegylated liposomal doxorubicin in ovarian cancer. *J. Ovarian Res. 15(1)*, 96. https://doi.org/10.1186/s13048-022-01029-6

Linders, A.N., Dias, I.B., López Fernández, T., Tocchetti, C.G., Bomer, N. & Van der Meer, P. (2024). A review of the pathophysiological mechanisms of doxorubicin-induced cardiotoxicity and aging. *npj Aging 10(1)*, 9. https://doi.org/10.1038/s41514-024-00135-7

Liu, Y., Chen, S., Wen, Z., Meng, J., Yang, Y., Zhang, Y., Wang, J. & Cao, X. (2024). Comparative pharmacokinetics of free doxorubicin and a liposomal formulation in cats following intravenous administration. *Front. Vet. Sci. 11*, 1353775. https://doi.org/10.3389/fvets.2024.1353775

Mattioli, R., Ilari, A., Colotti, B., Mosca, L., Fazi, F. & Colotti, G. (2023). Doxorubicin and other anthracyclines in cancers: activity, chemoresistance and its overcoming. *Mol Aspects Med. 93,* 101205. https://doi.org/10.1016/j.mam.2023.101205

Pernas, L. & Scorrano, L. (2016). Mito-morphosis: mitochondrial fusion, fission, and cristae remodeling as key mediators of cellular function. *Annu Rev Physiol. 78* 505-531. https://doi.org/10.1146/annurev-physiol-021115-105011

Prasanna, P.L., Renu, K. & Valsala Gopalakrishnan, A. (2020). New molecular and biochemical insights of doxorubicin-induced hepatotoxicity. *Life Sci. 250*, 117599. https://doi.org/10.1016/j.lfs.2020.117599

Rawat, P.S., Jaiswal, A., Khurana, A., Bhatti, J.S. & Navik, U. (2021). Doxorubicin-induced cardiotoxicity: An update on the molecular mechanism and novel therapeutic strategies for effective management. *Biomed. Pharmacother. 2021*, 111708. https://doi.org/10.1016/j.biopha.2021.111708

Shandilya, M., Sharma, S., Das, P.P. & Charak, S. (2020). Molecular-level understanding of the anticancer action mechanism of anthracyclines, in: Arnouk, H., Hassan, B.A.R. (Eds.), Advances in precision medicine oncology. *IntechOpen*, *9*. https://doi.org/10.5772/intechopen.94180

Shi, S., Chen, Y., Luo, Z., Nie, G. & Dai, Y. (2023). Role of oxidative stress and inflammation-related signaling pathways in doxorubicin-induced cardiomyopathy. *Cell Commun. Signal. 21(1)*, 61. https://doi.org/10.1186/s12964-023-01077-5

Sliwa, K., van der Meer, P., Petrie, M.C., Frogoudaki, A., Johnson, M.R., Hilfiker-Kleiner, D., Hamdan, R., Jackson, A.M., Ibrahim, B., Mbakwem, A., Tschöpe, C., Regitz-Zagrosek, V., Omerovic, E., Roos-Hesselink, J., Gatzoulis, M., Tutarel, O., Price, S., Heymans, S., Coats, A.J.S., Müller, C., Chioncel, O., Thum, T., de Boer, R.A., Jankowska, E., Ponikowski, P., Lyon, A.R., Rosano, G., Seferovic, P.M. & Bauersachs, J. (2021). Risk stratification and management of women with cardiomyopathy/heart failure planning pregnancy or presenting during/after pregnancy: a position statement from the Heart Failure Association of the European Society of Cardiology Study Group on peripartum cardiomyopathy. *Eur. J. Heart Fail. 23(4)*, 527-540. https://doi.org/10.1002/ejhf.2133

Sozen, E. & Ozer, N.K. (2017). Impact of high cholesterol and endoplasmic reticulum stress on metabolic diseases: an updated mini-review. *Redox Biol. 12*, 456-461. https://doi.org/10.1016/j.redox.2017.02.025

Szalay, C.I., Erdélyi, K., Kökény, G., Lajtár, E., Godó, M., Révész, C., Kaucsár, T., Kiss, N., Sárközy, M., Csont, T., Krenács, T., Szénási, G., Pacher, P. & Hamar, P. (2015). Oxidative/nitrative stress and inflammation drive progression of doxorubicin-induced renal fibrosis in rats as revealed by comparing a normal and a fibrosis-resistant rat strain. *PLoS One 10(6)*, e0127090. https://doi.org/10.1371/journal.pone.0127090

Timm, K.N., Ball, V., Miller, J.J., Savic, D., West, J.A., Griffin, J.L. & Tyler, D.J. (2022). Metabolic Effects of doxorubicin on the rat liver assessed with hyperpolarized MRI and metabolomics. *Front. Physiol. 12*, 782745. https://doi.org/10.3389/fphys.2021.782745

Townsend, N., Kazakiewicz, D., Lucy Wright, F., Timmis, A., Huculeci, R., Torbica, A., Gale, C.P., Achenbach, S., Weidinger, F. & Vardas, P. (2022). Epidemiology of cardiovascular disease in Europe. *Nat. Rev. Cardiol. 19(2)*, 133-143. https://doi.org/10.1038/s41569-021-00607-3

van den Boogaard, W.M.C., Komninos, D.S.J. & Vermeij, W.P. (2022). Chemotherapy side-effects: not all DNA damage is equal. *Cancers (Basel). 14(3)*, 627. https://doi.org/10.3390/cancers14030627

van der Zanden, S.Y., Qiao, X. & Neefjes, J. (2021). New insights into the activities and toxicities of the old anticancer drug doxorubicin. *FEBS J. 288(21)*, 6095-6111. https://doi.org/10.1111/febs.15583

Vitale, R., Marzocco, S. & Popolo, A. (2024). Role of oxidative stress and inflammation in doxorubicin-induced cardiotoxicity: a brief account. *Int. J. Mol. Sci. 25(13)*, 7477. https://doi.org/10.3390/ijms25137477

WHO (World Health Organization) (2023). World Health Organization model list of essential medicines – 23rd list, 2023, in: The selection and use of essential medicines 2023. web annex A. https://www.who.int/publications/i/item/WHO-MHP-HPS-EML-2023.02

Xie, S., Sun, Y., Zhao, X., Xiao, Y., Zhou, F., Lin, L., Wang, W., Lin, B., Wang, Z., Fang, Z., Wang, L. & Zhang, Y. (2024). An update of the molecular mechanisms underlying anthracycline induced cardiotoxicity. *Front. Pharmacol. 15, 1406247*. https://doi.org/10.3389/fphar.2024.1406247

Yang, F., Teves, S.S., Kemp, C.J. & Henikoff, S. (2014). Doxorubicin, DNA torsion, and chromatin dynamics. *Biochim Biophys. Acta. Rev. Cancer 1845(1)*, 84-89. https://doi.org/10.1016/j.bbcan.2013.12.002

Zhou, L., Shen, H., Li, X. & Wang, H. (2022). Endoplasmic reticulum stress in innate immune cells - a significant contribution to non-alcoholic fatty liver disease. *Front. Immunol. 13*, 951406. https://doi.org/10.3389/fimmu.2022.951406

STUDIA UNIVERSITATIS BABEŞ-BOLYAI BIOLOGIA, 71, 1, 2026 (p. 59-70)
doi:10.24193/subbbiol.2026.1.04

Functional study of PAP3/pTAC10 in the plastid-encoded RNA polymerase during chloroplast biogenesis

Leonard Calistru[1]✉, Florent Velay[2], Dorina Podar[1], and Robert Blanvillain[2]✉

[1]Babeş-Bolyai University; Str. Republicii nr. 44, Cluj-Napoca, Roumania; [2]CNRS, CEA, INRA, IRIG-LPCV Univ. Grenoble-Alpes Grenoble France;
✉***Corresponding authors, E-mail: leonard.calistru@gmail.com; robert.blanvillain@cea.fr***

Article history: Received 30 July 2025; Revised 24 February 2026; Accepted 30 April 2026; Available online 25 June 2026

Abstract. Chloroplast biogenesis in angiosperms depends on two types of RNA polymerases: the nuclear-encoded RNA polymerase (NEP) and the plastid-encoded RNA polymerase (PEP). PEP, in its active form (PEP-A), is supported by multiple nuclear-encoded proteins called PEP-associated proteins (PAPs). Among these, PAP3 (also known as pTAC10) plays a structural role in the PEP complex. Epifluorescent microscopy was used in order to verify its localization, confirming that PAP3 has a single subcellular localization: the chloroplast. Using the *Arabidopsis thaliana pap3-1* mutant and PHYB-GFP (PBG) reporter lines, we demonstrate that PAP3 is not required for PHYB-mediated light signaling pathway nor for photobody formation. Contrary to PAP8, whose loss of function affects red light signaling and photobody formation, *pap3* mutants show normal hypocotyl de-etiolation and photobody assembly under red light. These results suggest that in contrast with the nucleo-chloroplastic PAP8, PAP3 is confined to the plastid compartment and that it contributes exclusively to the plastidial transcriptional machinery. The difference in behavior between *pap8* and *pap3* mutants implements the existence of two distinct albino syndromes in PAPs, depending on their localization: dually localized (nucleus and chloroplasts) or solely localized (chloroplasts).

Keywords: chloroplast biogenesis, PAP3, photobodies, phytochrome B, plastid-encoded RNA polymerase.

Introduction

Plastids are cell organelles that are at the main sites of photosynthesis in eukaryotic cells. Their origin traces back to an event in which a cyanobacteria-like photosynthetic prokaryote was engulfed by a mitochondriated eukaryotic cell, resulting in a stable endosymbiosis of mutual benefit (Archibald, 2015). Within a multicellular photosynthetic organism, plastids can be divided based on their color, morphology, and ultrastructure into several types, such as: proplastids, chromoplasts or chloroplasts (Choi *et al.*, 2021).

The chlroplast has a much-reduced genome compared to its cyanobacterial ancestors because most of its genes were transfered to the nucleus of the eukaryotic cell following the primary endosymbiosis (Bock and Timmis, 2008). In order to accommodate these changes, plastid genes are transcribed by two types of RNA polymerases in angiosperms: the bacterial type plastid-encoded RNA polymerase (PEP) and one (RPOTp in monocots) or two (RPOTp and RPOTmp in dicots) nuclear-encoded RNA polymerase(s) (NEP) (Börner *et al.*, 2015). This division of labor between NEP and PEP allowed plants to regulate efficiently plastid gene expression under diverse conditions. While PEP transcribes photosynthesis-related genes, NEP alone transcribes a few housekeeping genes (Börner *a.*, 2015).

PEP exists in two forms: PEP-A and PEP-B. PEP-B is the low-active prokaryotic–like catalytic core enzyme composed of the α, α, β, β′, and β′′ subunits encoded by the plastid genes rpoA, rpoB, rpoC1, and rpoC2. The PEP-A complex consists of the catalytic core enzyme associated with nuclear-encoded additional proteins named PAPs (Steiner *et al.*, 2011; Liebers *et al.*, 2022).

PEP-associated proteins (PAPs) are structurally and functionally diverse, and can be divided based on their potential function, such as DNA/RNA metabolism, redox regulations, ROS protection, or unknown (Kindgren and Strand, 2015). Functional inactivation via T-DNA insertion in most PAP genes lead to an arrest in chloroplast biogenesis, resulting in ivory or albino plants that are only viable upon the addition of a carbon source (e.g. sucrose) in sterile growth conditions (Liebers *et al.*, 2022). Out of the 12 PAPs identified so far, six contain a predicted nuclear localization signal (NLS) according to the algorithm of cNLS mapper. Some of these PAPs, such as PAP8 or PAP5, display a nucleo-chloroplastic accumulation of the GFP-tagged proteins, meaning that they are dually localized to the nucleus and chloroplast (Liebers *et al.*, 2022).

In angiosperms, chloroplast development, perceived as an initial greening, is preceded by light activation of photoreceptors that trigger photomorphogenesis. One key photoreceptor is phytochrome-B who senses red (600-700 nm) and far-red light (700-750 nm), and triggers the transition from skotomorphogenesis (growth in the dark) to photomorphogenesis (growth in the light), promoting

the development of green chloroplasts, de-etiolation and inhibition of stem elongation (Liebers *et al.*, 2020). Phytochrome-B, alongside other phytochromes (A to E), is found in membraneless organelles called photobodies (Yoo *et al.*, 2019). Initially appearing as small, transient foci, photobodies mature into larger, more stable structures under continuous red light exposure, suggesting a functional transition from early signaling platforms to photostable repositories for active phytochromes (Huang *et al.*, 2016). Their composition includes not only phytochromes but also key transcriptional regulators like PAP5/HEMERA/pTAC12, which is required for photobody formation and photomorphogenesis (Chen *et al.*, 2010; Yoo *et al.*, 2019).

Here, we show first that PAP3 fused to a fluorescent tag is solely localized in plastids and second that the sporophytic-lethal mutant *pap3* display a good response to over-accumulation of the phytochrome-B-GFP fusion protein. We then propose to distinguish the existence of two albino syndromes within *paps* albino mutants according to the presence or absence of a predicted NLS within the coding sequence.

Materials and methods

Plant materials and growth conditions. *Arabidopsis thaliana* Columbia-0 (Col-0), the *pap3-1* T-DNA insertion mutant (CS16115), and the *pap3*-PHYB-GFP transgenic line were used. Seeds (~5000) were surface-sterilized in a 1.5 mL Eppendorf tube by two washes with 750 µL of 70% ethanol for 1 minute each, with gentle mixing by inversion and removal by pipetting. After ethanol removal, 60 µL of Domestos (commercial bleach, ~5% NaOCl) was added, and the seeds were mixed by tube inversion for 8 minutes. To preserve seed viability, total exposure to sterilizing agents did not exceed 20 minutes. The seeds were then washed three times with 1 mL of sterile water and placed on Murashige and Skoog (MS) medium, ensuring proper spacing. Seeds were then stratified at 4°C in darkness for 48 hours to synchronize germination. Growth conditions included two distinct light treatments: (1) continuous red light (660 nm/ ~ 8 $\mu mol.m^{-2}.s^{-1}$ photons) for 5 days and (2) continuous far-red light (730 nm) for 5 days.

Cellular lines. Transformation experiments utilized DH5α and One Shot™ TOP10 *Escherichia coli* competent cells stored at -80°C. The DH5α strain (F- Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rk-, mk+) phoA supE44 thi-1 gyrA96 relA1 λ-) and the TOP10 strain (F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 lacX74 recA1 ara139 Δ(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG) were used to maximize transformation efficiency.

Gentamicin-resistant pSOUP-containing electrocompetent *Agrobacterium tumefaciens* (strain C58 (rifR), Ti pMP90 (pTiC58DT-DNA) (gentR), pSOUP-p19 (tetR), nopaline) stored at -80°C was used for plant transformation.

Genetic materials: Vectors used included TOPO-BLUNT PCR cloning vectors (1 µL per topoisomerase reaction), binary vectors, and previously constructed plasmids (pQD11a, pCL02a9, pMT13, pBB330a*, pBB300, pBB301, pBB213, pCL23b, pKP20, pKP34). PCR-amplified inserts were obtained using specific primer pairs designed for cloning and verification purposes.

Plasmid DNA extraction: Plasmid DNA was extracted and purified using the ZymoPURE Plasmid Miniprep Kit. The procedure involved bacterial cell resuspension, alkaline lysis, and neutralization. Plasmid DNA was selectively bound to a silica matrix under high-salt conditions, followed by sequential washes to remove proteins, RNA, and other contaminants. Purified DNA was eluted in nuclease-free water and stored at -20°C.

DNA quantification: DNA concentration and purity were assessed using a NanoDrop™ 2000 spectrophotometer (Thermo Scientific). One microliter of nuclease-free water was used as a blank for baseline correction. DNA absorbance was measured at 260 nm, and purity was evaluated using the A260/A280 and A260/A230 ratios, where values >2 indicated minimal protein or RNA contamination.

Restriction enzyme digestion: Plasmid and PCR product digestions were performed using XhoI and BamHI for cloning, and NcoI, XbaI, HindIII, SacI, and SpeI for verification. Each reaction contained 2 µL of template DNA, 2 µL of 10X rCutSmart buffer, and the enzymes were used according to their unit definition on the concentration of vector and water in a total volume of 20 µL. Digestion was carried out at 37°C for 1 hour.

***Escherichia coli* and *Agrobacterium tumefaciens* transformation**. For *Escherichia coli* transformation, competent cells were prepared using 5xKMC buffer (500 mM KCl, 150 mM $CaCl_2$, 250 mM $MgCl_2$) prepared in advance, filter-sterilized (0.22 µm), and stored at -20°C. Transformation was performed by mixing 20 µl of 5xKMC, 70 µl of sterile water, and 10 µl of plasmid DNA. The mixture was added to the bacterial suspension and incubated on ice for 20 minutes, followed by a 10-minute incubation at room temperature. One milliliter of prewarmed LB or SOC medium was added, and the culture was incubated at 37°C for 1 hour with shaking. Transformed bacteria were plated on LB agar

supplemented with antibiotic and incubated overnight at 37°C. For *A. tumefaciens*, the transformation is done by electroporation because it is more efficient on large plasmids and Gram positive cells, such as *A. tumefaciens*. Five µL of vector were added to 50 µL of *A. tumefaciens* competent cells. These latter were then electrically shocked at 2500 V for 5 ms. Afterwards, the colonies were placed on LB agar medium with antibiotics such as: kanamycin, carbenicillin or spectinomycin. The Petri dishes were left overnight at 28°C under agitation and checked the next 2 days to assess the presence of resistant colonies.

Electrophoresis on agarose gel. The 1% agarose gel was prepared by adding 3 g of Roti®agarose to 300 mL of TAE 1X then warming while mixing periodically. For one gel, 25 mL of this solution was used alongside 2 µL of Red™ Safe 10,000X which was added in order to solidify the solution into a gel. Red™ Safe is a DNA intercalating agent which is added to allow DNA visualization under UV light emitted from the GelDoc™ BioRad. The solution was then poured into a chamber. After 20-30 minutes, the gel was immersed in TRIS Acetate EDTA 1% buffer, then the samples were injected with a loading dye (6X) into the wells. Thermo Scientific GeneRuler DNA Ladder Mix (5 µL) was also added into the wells to quantify the size of each band. The 1% agarose gel was submitted to an electric field at 110 V for 25 minutes in a solution of TAE 1X. After migration, it was analyzed under UV light (300 nm) with the Gel Doc™ BioRad.

Ligation. Ligation reactions were performed using a 3:1 molar ratio of insert to vector. This calculation was done with the following formula: (DNA mass/molecular weight)x$\mathcal{N}_{\text{Avogadro}}$. DNA mass is determined by comparison to known quantities of the ladder after an electrophoresis (Figure 6.b), the molecular weight is the number of base pairs multiplied by the weight of one base pair (660 g/mole), and Avogadro's number is a constant: 6.022×10^{23}. After determining the quantities of insert/vector that were being used, T4 DNA ligase was added with T4 buffer containing ATP in order to synthesize the phosphodiester bound. The samples were then left at room temperature for 4 hours.

Biolistic on onions. Also known as "particle bombardment", is a genetic transformation method done on plant cells where DNA is introduced into plant tissue, in our case onion epidermis, for transient expression. The process was the following:

- DNA mixture preparation: ~2,5 µg of gold beads in a final volume of 20 µL.
- Samples preparation with gold: 20 µL of Binding Buffer (Seashell™) is added into the DNA mixture. Preparation of 30 µL of gold (for 3 probes). 40 µL of Binding Buffer DNA mixed on the gold. Resuspended well by

pipetting and vortexing. After being kept on ice for 5 min, 70 μL of Precipitation Buffer is added. The supernatant is discarded after the suspension has been vortexed and kept on ice for 5-10 min. 500 μL of absolute Ethanol (100%) is added. Afterwards, the ethanol was removed and 30 μL of Ethanol 100% were added for pipetting the gold particles.

- Onion preparation: Fresh and turgid yellow onions were selected. Pieces were cut in such a way that the sample obtained had an intact epidermis. The annotation was done on the external face of it.
- Shooting: First, the pump has been turned on for vacuum, helium and PDS1000He then all of the required elements were placed (rupture disk, macro carrier, micro carrier, and stop grid). Furthermore, the vacuum switch was pressed until the gauge on the top of the gun read - 27 mm Hg, then the switch was put on "Hold". After that, the "Fire" switch was pressed and held until it popped at the rupture pressure. Finally, the vacuum was vented and the door opened. The onions were kept overnight in the dark before observation.

Primers. Primer sequences were designed using ApE (A Plasmid Editor; https://jorgensen.biology.utah.edu/wayned/ape), and used to generate PAP3 coding sequence fragments carrying silent point mutations that disrupt internal restriction enzyme sites without altering the encoded aminoacids. Those primers were also used to verify the presence of the modified PAP3 constructs in bacterial colonies via PCR.

Sequencing. To confirm the presence of the desired point mutations in the PAP3 coding sequence, plasmid DNA (15 μL at 100 ng·μL^{-1}) together with the corresponding primers was submitted to Eurofins Genomics (Ebersberg, Germany) for Sanger sequencing.

Results

Cloning PAP3

XhoI and BamHI are two restriction sites that were added using PCR at the beginning and end of the PAP3 coding sequence, because they are suitable for our cloning strategies. The problem is the presence of one more of each inside the coding sequence, meaning that some modifications had to be made in order to use available plasmids in the lab (Fig. 1A). Two point mutations were design so that the amino acid produced will not be changed, but the restriction sites would be disrupted. In the XhoI site, CTC was changed to CTT, both codons

encoding leucine, while in the BamHI site, ATC was changed to ATT, both encoding isoleucine (Fig. 1A. Those changes were made with specific primers, design in such a way that they would align to the template DNA even with those point mutations thanks to the high number of cytosine (C) and guanine (G) nucleotides at both ends.

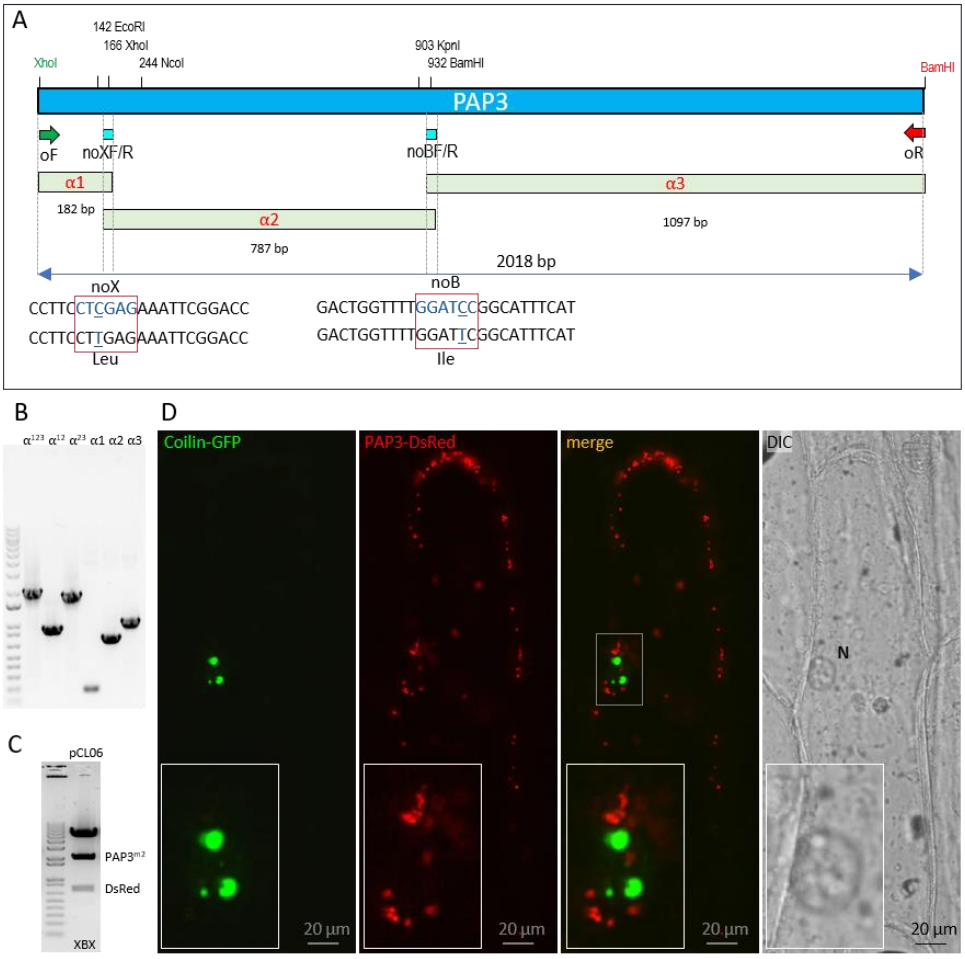


Figure 1. Cloning strategy and localization of PAP3 in the chloroplast. A. Map of the PAP3 coding sequence including restriction sites, the three α fragments derived from the PAP3 sequence, and location of primers. B. Phusion PCR and fragment extension PCR of α fragments. C. Enzyme digestion of pCL06 with XhoI/BamHI/XbaI. D. Subcellular localization of PAP3-DsRed compared to nuclear marker coilin-GFP in onion epidermal cells following biolistic transformation.

The PAP3 sequence was divided in three parts (α1, α2 and α3) in order to ensure a proper product with no XhoI/BamHI restriction sites within (Fig. 1B). The fragments were obtained with the corresponding primers in a Hot Start Phire PCR, and the final recombinant sequence was achieved via fragment extension PCR (Fig. 1B).

Red Fluorescent Protein (DsRed2) is a fluorescent marker that was added to PAP3 via cloning in order to continue the experiments in various directions. The resulting construct was verified with restriction enzymes for the presence of PAP3 and DsRed (Fig. 1C).

Localization of PAP3 in the chloroplast

Subcellular localization of PAP3 was done with biolistic transformation on epidermal onion cells using PAP3-DsRed (pCL06). Microscopy was employed to capture high-resolution fluorescence signals across multiple channels. Coilin-GFP fluorescent signal is restricted to the nucleus, confirming the nuclear compartment's spatial integrity (Fig. 1D). DsRed channel reveals discrete fluorescence exclusively within plastid-like structures, indicating robust localization of PAP3 to the chloroplast (Fig. 1D). Notably, no DsRed signal was observed in the nucleus. Merged image of both channels (Fig. 1D), highlighting the clear spatial segregation between PAP3-DsRed and coilin-GFP signals, and further supporting the absence of nuclear targeting for PAP3. Differential Interference Contrast image (DIC) delineates cellular boundaries and validates the structural context of the fluorescence signals (Fig. 1D).

PHYB-mediated light response

Stable overexpression of phytochrome PHYB-GFP (PBG) is known to mediate hypersensitivity of *Arabidopsis* seedlings to red light leading to a significant inhibition of hypocotyl elongation when compared to wild-type (Fig. 2A, 2B). After introducing PBG into *pap3* mutant background, this PBG effect remained the same, meaning that PAP3, as opposed to PAP8, does not play a role in the PHYB-mediated light response (Fig. 2B, 2C, 2D). This finding serves as evidence towards the difference in albino syndromes between PAPs with a single subcellular localization (PAP3 in chloroplast) and ones with dual localization (PAP8 in nucleus and chloroplast).

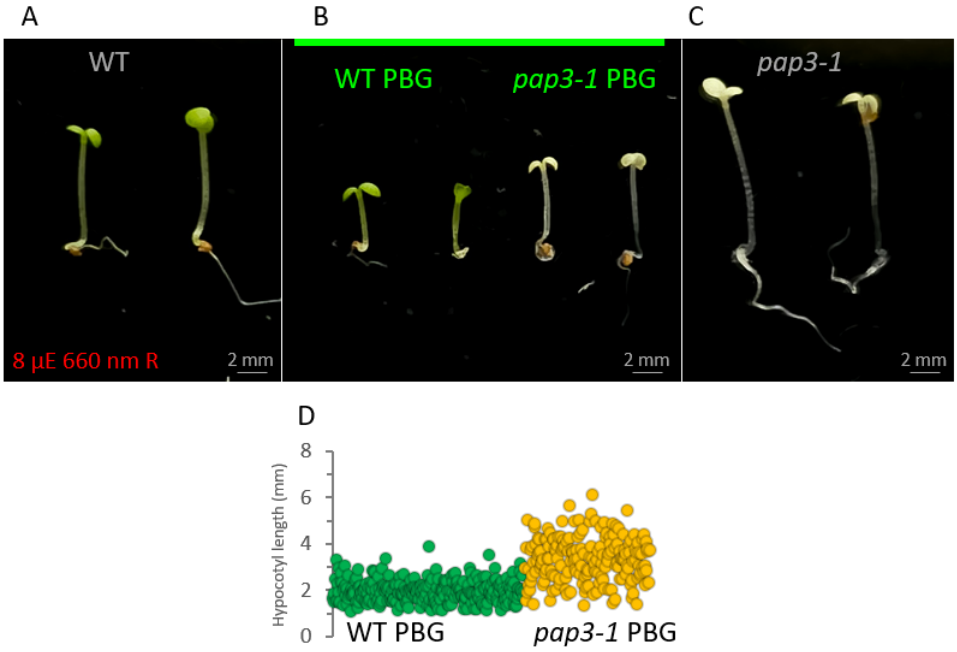


Figure 2. PAP3 is not essential for the PHYB-mediated light induction of photomorphogenesis. A-C. Phenotypes of given genotypes subjected to 5 days of illumination at 8 µmol m^{-2} s^{-1} 660-nm red light. D. Data analysis comparing the hypocotyl size between wild type and *pap3-1* mutant.

Photobody formation

Photobodies are dynamic subnuclear structures where activated phytochromes, particularly PHYB, accumulate in response to light. In wild-type *Arabidopsis thaliana* seedlings, active PHYB rapidly forms large, stable photobodies upon red light exposure, which correlates with robust transcriptional reprogramming and repression of hypocotyl elongation (Fig. 3A. Conversely, in *pap8* mutants, PHYB fails to form these large photobodies efficiently, remaining only small and numerous early photobodies. In contrast, PAP3 does not appear to share this functionality. In the *pap3* mutant background, photobody formation proceeds normally in the absence of PAP3 (Fig. 3B). The shape, number and dimensions are relatively the same when comparing *pap3* mutant to the WT, and the transition from early, small and numerous photobodies to late, big and few in number is not altered at all between the two, further distinguishing PAP3 from PAP8 in this context (Fig. 3B, 3C).

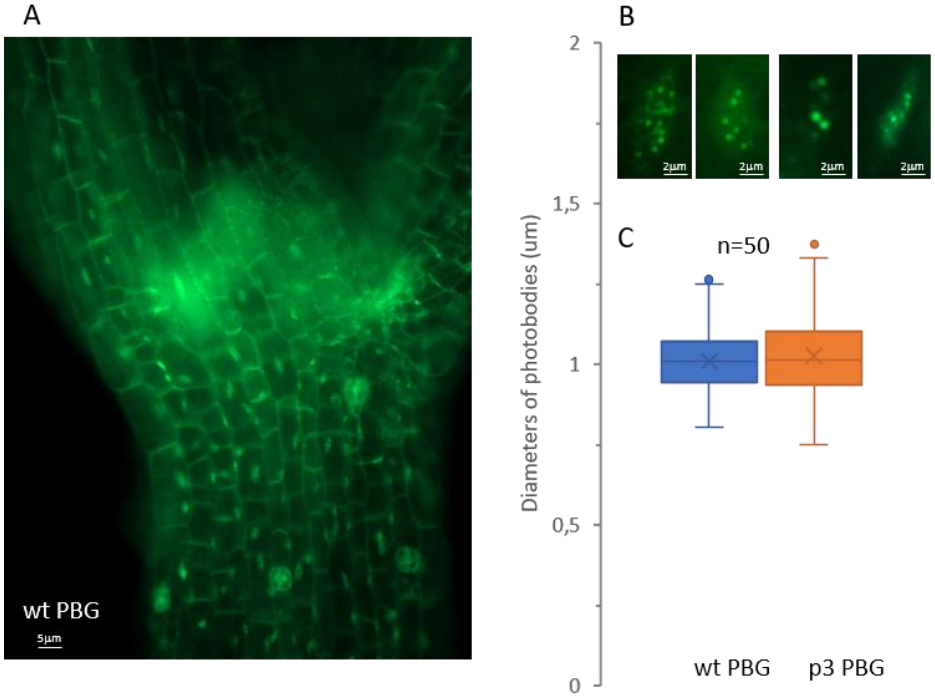


Figure 3. Photobody formation is not hindered in the *pap3-1* mutant. A. Whole-seedling image showing general GFP signal distribution across the plant. B. Accumulation of PBG observed under GFP excitation in the given genotypes (wt-PBG/*pap3-1* PBG) using epi-fluorescence microscopy. C. Data analysis of phodobodies size from the PBG and *pap3-1* PBG lines.

Discussion

Plastid gene expression in angiosperms relies on the plastid-encoded RNA polymerase (PEP) complex and its associated nuclear-encoded proteins called PAPs, several of which have been involved in plastid transcription and nuclear light signaling pathways. Previous studies showed that some PAPs, such as PAP8/pTAC6 and PAP5, display dual localization to plastids and the nucleus and play a role in coordinating chloroplast biogenesis with photomorphogenesis (Liebers *et al*., 2020; Liebers *et al*., 2022).

In this study, it is shown that PAP3/pTAC10 is exclusively localized to plastids and does not participate in PHYB-mediated light signaling pathway. Subcellular localization experiments revealed no detectable nuclear accumulation

of PAP3, in contrast to dually localized PAPs previously described in literature. This plastid-confined localization is consistent with earlier proteomic studies identifying PAP3 as a structural component of the transcriptionally active plastid chromosome (pTAC), rather than as a signaling factor involved in nucleus–plastid communication (Pfannschmidt *et al*., 2000; Steiner *et al*., 2011).

Functional analyses further support this distinction. While *pap8-1; PBG* exhibit defects in hypocotyl de-etiolation and impaired photobody formation under red light, *pap3-1; PBG* display normal PHYB nuclear body assembly and red light hypersensitivity. These observations indicate that PAP3 is not one of the main actors in the PHYB-mediated light signaliing pathway, reinforcing the idea that its function is confined to the plastid compartment (Kindgren and Strand, 2015).

Together, these results support a functional distinction among PAPs based on their subcellular localization. Dually localized PAPs have been shown to play a role in both plastid gene expression and nuclear light-regulated processes, whereas plastid-confined PAPs, such as PAP3, appear to function primarily within the plastid transcriptional machinery. Therefore, the absence of PAP3 does not impair PHYB-mediated light signaling pathway, indicating that its function is not required for photobody formation or red light–dependent photomorphogenic responses.

This distinction refines current models of plastid–nucleus communication by emphasizing that not all components of the PEP complex contribute to retrograde signaling. Instead, the regulatory scope of individual PAPs appears to be tightly linked to their subcellular localization and molecular function within the PEP complex.

Conclusions

This study identifies PAP3/pTAC10 as a plastid-confined component of the plastid-encoded RNA polymerase complex. Fluorescent localization analyses demonstrate that PAP3 localizes exclusively to plastids and does not enter the nucleus.

Functional characterization of the *Arabidopsis pap3-1* mutant shows that PAP3 is not required for PHYB-mediated light signaling pathway, hypocotyl de-etiolation, or photobody formation. These findings distinguish PAP3 from dually localized PAPs, such as PAP8, which are involved in both plastid transcription and nuclear light signaling.

Overall, these results support a functional classification of PEP-associated proteins based on subcellular localization, in which plastid-confined PAPs primarily affect chloroplast biogenesis, whereas dually localized PAPs integrate plastid development with nuclear photomorphogenic pathways.

References

Archibald, J. M. (2015). Endosymbiosis and eukaryotic cell evolution. *Current Biology*, *25*(19), R911–R921. https://doi.org/10.1016/j.cub.2015.07.055

Bock, R., & Timmis, J. N. (2008). Reconstructing evolution: Gene transfer from plastids to the nucleus. *BioEssays*, *30*(6), 556–566. https://doi.org/10.1002/bies.20761

Börner, T., Aleynikova, A. Y., Zubo, Y. O., & Kusnetsov, V. V. (2015). Chloroplast RNA polymerases: Role in chloroplast biogenesis. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1847(9)*, 761–769. https://doi.org/10.1016/j.bbabio.2015.02.004

Cappadocia, L., Maréchal, A., Parent, J.-S., Lepage, étienne, Sygusch, J., & Brisson, N. (2010). Crystal Structures of DNA-Whirly Complexes and Their Role in *Arabidopsis* Organelle Genome Repair. *The Plant Cell*, *22*(6), 1849–1867. https://doi.org/10.1105/tpc.109.071399

Chan, C. X. & Bhattacharya, D*.* (2010*). The Origin of Plastids. Nature Education 3(9):84.*

Chen, M., Galvão, R. M., Li, M., Burger, B., Bugea, J., Bolado, J., & Chory, J. (2010). *Arabidopsis* HEMERA/pTAC12 initiates photomorphogenesis by phytochromes. *Cell*, *141*(7), 1230–1240. https://doi.org/10.1016/j.cell.2010.05.007

Chen, M., Schwab, R., & Chory, J. (2003). Characterization of the requirements for localization of phytochrome B to nuclear bodies. *Proceedings of the National Academy of Sciences*, *100*(24), 14493–14498. https://doi.org/10.1073/pnas.1935989100

Choi, H., Yi, T., & Ha, S.-H. (2021). Diversity of plastid types and their interconversions. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/fpls.2021.692024

Huang, H., Yoo, C. Y., Bindbeutel, R., Goldsworthy, J., Tielking, A., Alvarez, S., Naldrett, M. J., Evans, B. S., Chen, M., & Nusinow, D. A. (2016). PCH1 integrates circadian and light-signaling pathways to control photoperiod-responsive growth in *Arabidopsis*. *eLife*, *5*, e13292. https://doi.org/10.7554/eLife.13292

Kindgren, P., & Strand, Å. (2015). Chloroplast transcription, untangling the Gordian Knot. *New Phytologist*, *206*(3), 889–891. https://doi.org/10.1111/nph.13388

Liebers, M., Cozzi, C., Uecker, F., Chambon, L., Blanvillain, R., & Pfannschmidt, T. (2022). Biogenic signals from plastids and their role in chloroplast development. *Journal of Experimental Botany*, *73*(21), 7105–7125. https://doi.org/10.1093/jxb/erac344

Liebers, M., Gillet, F., Israel, A., Pounot, K., Chambon, L., Chieb, M., Chevalier, F., Ruedas, R., Favier, A., Gans, P., Boeri Erba, E., Cobessi, D., Pfannschmidt, T., & Blanvillain, R. (2020). Nucleo-plastidic PAP8/pTAC6 couples chloroplast formation with photomorphogenesis. *The EMBO Journal*, *39*(22), e104941. https://doi.org/10.15252/embj.2020104941

Pfannschmidt, T., Ogrzewalla, K., Baginsky, S., Sickmann, A., Meyer, H. E., & Link, G. (2000). The multisubunit chloroplast RNA polymerase A from mustard (*Sinapis alba*). *European Journal of Biochemistry*, *267*(1), 253–261. https://doi.org/10.1046/j.1432-1327.2000.00991.x

Steiner, S., Schröter, Y., Pfalz, J., & Pfannschmidt, T. (2011). Identification of essential subunits in the plastid-encoded RNA polymerase complex reveals building blocks for proper plastid development1[C][W][OA]. *Plant Physiology*, *157*(3), 1043–1055. https://doi.org/10.1104/pp.111.184515

Yoo, C. Y., Williams, D., & Chen, M. (2019). Quantitative analysis of photobodies. In A. Hiltbrunner (Ed.), *Phytochromes* (Vol. 2026, pp. 135–141). Springer New York. https://doi.org/10.1007/978-1-4939-9612-4_10

STUDIA UNIVERSITATIS BABEŞ-BOLYAI BIOLOGIA, 71, 1, 2026 (p. 71-100)
doi:10.24193/subbbiol.2026.1.05

Leveraging the power of computational immunology to develop a novel couplet mRNA vaccine concatenating conserved epitopes of monkeypox virus antigens

Leana Rich Herrera-Ong

Department of Biochemistry and Molecular Biology, College of Medicine, University of the Philippines Manila, Manila City, Philippines
E-mail: lmherrera2@up.edu.ph

Article history: Received 9 August 2025; Revised 2 December 2025; Accepted 16 March 2026; Available online 25 June 2026

Abstract. The monkeypox (Mpox) disease was declared as a Public Health Emergency of International Concern (PHEIC) in 2022 and in 2024. Currently, there are no Mpox-specific vaccines and antiviral drugs for the treatment of Mpox. This study developed a novel couplet mRNA vaccine through immunoinformatics tools and databases. Highly conserved sequences of nine Mpox antigens were identified and utilized for T-cell and B-cell epitope mapping. Using data obtained from experimentally validated epitopes, the potential immunogenicity of predicted T-cell epitopes was evaluated. Peptides that may cause cross-reactivity, toxicity, and allergic reactions were excluded. The population coverage of candidate epitopes was also estimated. Epitopes were arranged and adjoined to form the mRNA constructs, aiming to direct its processing either to exogenous or endogenous pathways, while maximizing cleavage scores to produce the desired peptides. Conservancy analysis revealed that the nine Mpox target antigens A29L, A35R, A42R, B6R, E8L, F13L, H3L, L1R, and M1R are highly conserved. After the exclusion of potentially harmful and non-immunogenic peptides, 559 CD4+ and 104 CD8+ T-cell epitopes remained. The couplet mRNA constructs encode potentially immunogenic T-cell epitopes and accessible B-cell epitopes of multiple Mpox antigens that can be expressed and processed to generate the desired peptide sequences in humans. Estimated population coverage for mTcPox and mBThPox

constructs are 77.67% and 81.81%, respectively. This study developed a novel couplet mRNA Mpox vaccine design containing highly conserved, immunogenic epitopes, with acceptable safety profile. Further *in vitro*, animal, and clinical studies are anticipated for human applications.

Keywords: epitope; immunoinformatics; monkeypox; Mpox; mRNA vaccines

Introduction

The World Health Organization (WHO) declared the monkeypox (Mpox) virus outbreak—a zoonotic disease—as a Public Health Emergency of International Concern (PHEIC) twice: first in 2022 and again in 2024 (Sagdat *et al*., 2024). As of 2025, the global prevalence of Mpox continues to rise, particularly in African countries, with imported cases increasingly reported in other regions. Individuals with compromised immune systems and those with multiple sexual partners face a heightened risk of infection. Between January 2022 and March 2025, a total of 137,892 confirmed cases and 317 deaths were reported across 132 countries, with the majority of fatalities occurring in African nations (Yadav *et al*., 2025).

Two major clades of the Mpox virus—Clade I and Clade II—are currently circulating worldwide. Clade II has been primarily responsible for the widespread outbreak from 2022 to 2025. More recently, the emergence of Mpox Clade Ib in Southeast Asia (SEA) has raised concerns, especially for countries already grappling with Clade IIb since 2022 (A *et al*., 2025). The mortality rate of Mpox varies depending on the viral strain, ranging from 10% to 15% (Bogacka *et al*., 2025).

The natural reservoirs of the Mpox virus are believed to be primates and certain rodent species native to Africa. Human-to-human transmission occurs through direct contact with mucous membranes, body fluids, tissues, or respiratory droplets from infected individuals (Beeson *et al*., 2023). The severity of Mpox infection is influenced by the patient's immune status, the specific viral strain involved, and the presence of complications. The virus typically enters the body through skin lesions or mucous membranes and then disseminates systemically. The latent period lasts up to two weeks, during which infected individuals remain asymptomatic and show no visible lesions. Following this phase, symptoms such as fever, muscle pain, headache, and lymphadenopathy emerge and persist for approximately three days. Subsequently, a rash develops over a span of two to four weeks, progressing from papules to pustules and vesicles. These lesions eventually crust over and heal, often leaving scars (Zahmatyar *et al*., 2023).

Mpox virus belongs to the family Poxviridae and the genus *Orthopoxvirus*. It is characterized by its oval morphology and a dumbbell-shaped nucleocapsid encased in lipid-containing particles. The viral genome is approximately 197 kb in size and encodes around 190 proteins. The virus exists in two distinct forms: extracellular enveloped virions (EEV) and intracellular mature virions (IMV). IMVs have a single membrane and enter host cells via fusion and endocytosis, being released upon cell lysis. EEVs, which possess a double membrane, attach to host mucous membranes through glycosaminoglycans (GAGs) and enter via membrane fusion (Lu *et al*., 2023).

Targeting EEV surface proteins such as B6R and A35R may help reduce viral transmission. F13L, another EEV protein, is embedded within the viral envelope and is essential for the formation of extracellular viral particles (Callaby *et al*., 2025). IMV-associated antigens include A29R, H3L, E8L, L1R, and M1R. A29R facilitates viral fusion, replication, and egress. H3L enhances viral binding to host cells, increasing infectivity. E8L promotes surface attachment and entry, while L1R and M1R contribute to viral entry, assembly, and neutralization. A42R, a profilin-like protein, interacts with actin filaments, but its role in Mpox pathogenesis remains unclear (Sagdat *et al*., 2024). Taken together, these antigens play critical roles in Mpox virulence and represent promising targets for antiviral drug and vaccine development.

Smallpox vaccines approved for Mpox offer only partial efficacy, and viral transmission continues due to the lack of Mpox-specific vaccines or drugs. Current antivirals—tecovirimat, brincidofovir, and cidofovir—were originally developed for smallpox (Bogacka *et al*., 2025). Advancing the development of Mpox-targeted therapeutic agents is therefore essential. Immunoinformatics provides powerful tools for identifying and designing vaccine components (Rawat *et al*., 2023). This study aims to develop a novel couplet Mpox mRNA vaccine using advanced immunoinformatic techniques. The vaccine includes specific, conserved, and immunogenic B-cell and T-cell epitopes targeting multiple Mpox antigens. Each epitope is directed to either endogenous or exogenous pathways to optimize immune processing. The couplet mRNA sequences are designed for enhanced expression in human cells and improved molecular stability, making them suitable for future immunotherapeutic applications.

Materials and methods

Identification of conserved sequences in Mpox antigens

The protein sequences of Mpox antigens were obtained from the National Center for Biotechnology Information (NCBI) on June 5, 2025. The full lengths of A29L, A35R, A42R, B6R, E8L, F13L, H3L, L1R, and M1R are 110, 181, 133, 317, 304,

372, 324, 152 and 250 amino acid residues, respectively. Redundant sequences were removed in the Galaxy CD-HIT server (https://usegalaxy.eu/?tool_id=cd_hit). Variable positions were identified and masked in Protein Variability Server (http://imed.med.ucm.es/PVS/) using the Shannon variability threshold ≤ 0.1 and fragments with at least 6 residues.

Mapping of CD8+ and CD4+ epitopes

Cytotoxic T-cell epitopes were identified using the recommended algorithm in IEDB–NetMHCIPan 4.1 BA (https://nextgen-tools.iedb.org/pipeline?tool=tc1). The IC_{50} threshold used to classify good binders is < 500 nM (Reardon *et al*., 2021). To increase prediction accuracy, the thresholds for proteasome and TAP scores were set to > 1.0 (Olotu *et al*., 2021). Helper T-cell epitopes were mapped using IEDB NetMHCIIPan 4.1 BA, with IC_{50} < 500 nM to identify good binders. The Uniprot database (https://www.uniprot.org/) was consulted to determine known residues in the target antigens with post-translational modifications (PTM). Epitopes with PTM residues were excluded from the set of T-cell epitopes to be assessed in subsequent steps.

Assessment of potentially immunogenic helper and cytotoxic T-cell epitopes

The set of mapped cytotoxic T-cell epitopes were further filtered using three different state-of-the-art immunogenicity tools. DeepImmuno (DIm) uses a convolutional neural network (CNN) algorithm to predict the immunogenicity of MHC–peptide pairs (https://deepimmuno.research.cchmc.org/). DIm was found to have high average auROC (0.85) and auPR (0.81) (Li *et al*., 2021). The T cell class I p-MHC immunogenicity predictor (CIImm) in IEDB (http://tools.iedb.org/immunogenicity/) assesses immunogenicity by providing an immunogenicity score calculated as a position-dependent weighted sum of the non-anchor amino acids of the neoantigen amino acid sequence (Nibeyro *et al*., 2023). PRIME 2.0 is another state-of-the-art immunogenicity tool that combines affinity predictions with TCR-recognition propensity (http://prime.gfellerlab.org). Rank is provided and a non-immunogenic peptide receives a score of zero (Gfeller *et al*., 2023). Currently, there is no consensus threshold value to identify immunogenic epitopes. In general, the higher the immunogenicity score, the higher the probability of an epitope being immunogenic (Nibeyro *et al*., 2023). To create a more stringent standard in this study, the structures of five experimentally validated cytotoxic T-cell epitopes with their MHC I binders, were obtained and utilized as positive references for the immunogenicity analysis of predicted cytotoxic T-cell epitopes. These five experimentally validated CD8+ T-cell epitopes are:

HPV.16 E7 peptide LLMGTLGIV (6APN), HIV epitope SLYNTVATL (5NMD), SARS-CoV-2 spike-derived peptide YLQPRTFLL (7RTD), SARS-CoV-2 spike-derived peptide K417T mutant TIADYNYKL (7UM2), and SARS-CoV-2 spike-derived peptide YIWLGFIAGL (7UR1). The immunogenicity scores of experimentally validated CD8+ T-cell epitopes were determined in DIm, CIIm and PRIME 2.0. This study excluded all predicted cytotoxic T-cell epitopes with immunogenicity scores lower than any of the lowest immunogenicity scores of the 5 positive references.

The release of interferon-gamma (IFN-gamma) is a hallmark of Th1 response that is critical to mitigate intracellular pathogens, such as Mpox viruses. IFN-epitope server (http://crdd.osdd.net/raghava/ifnepitope/) was utilized to identify which among the predicted CD4+ T-cell epitopes can potentially induce IFN-γ. The study selected the hybrid algorithm which combines the advantages of SVM-based and motif-based techniques (Dhanda *et al.,* 2013). IL4pred server (https://webs.iiitd.edu.in/raghava/il4pred/predict.php) was utilized to identify helper T-cell epitopes that can potentially induce IL4. Using motif information and amino acid pairs, the accuracy of this method is found to be 75.76% (Nebangwa *et al.,* 2024). With an SVM threshold of 2.0, the hybrid SVM-based and motif-based technique was employed to identify IL4-inducing CD4+ epitopes. The remaining epitopes were further analyzed for potential cross-reactivity, toxicity, and allergenicity.

Mapping of B-cell epitopes in the extracellular enveloped virus proteins of Mpox

Mpox B6R and A35R are two most accessible extracellular enveloped virus (EEV) proteins that play pivotal roles in cell attachment and effective spreading. EEV is an enveloped form of virus released from the cell that is capable of infecting other cells. Thus, peptide sequences of B6R and A35R were retrieved to identify potential B-cell epitopes. ABCpred (https://webs.iiitd.edu.in/raghava/abcpred/index.html) tool, together with six other well-established linear B-cell epitope prediction tools (Bepipred Linear Epitope Prediction 2.0, Chou & Fasman Beta-Turn Prediction, Emini Surface Accessibility Prediction, Karplus & Schulz Flexibility Prediction, Kolaskar & Tongaonkar Antigenicity, and Parker Hydrophilicity Prediction) in IEDB (http://tools.iedb.org/bcell/) were applied to map B-cell epitopes using default thresholds. Using 6-kmer consensus sequences, peptide sequences that overlap in at least 5 different tools were included. Resulting epitopes that are located in the transmembrane and internal regions were removed. The transmembrane, internal and external domains of each sequence were identified using the Interproscan tool (https://www.ebi.ac.uk/interpro/search/sequence/). B-cell epitopes with variable residues and PTMs were also excluded. The remaining

B-cell epitopes were further evaluated for potentially cross-reactive, allergenic, and toxic peptides. Lastly, overlapping B-cell epitopes per antigen were integrated into a single peptide sequence.

Exclusion of potentially cross-reactive, toxic and allergenic T-cell and B-cell epitopes

All the epitopes were thoroughly inspected using the BLASTP tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins), to exclude peptides having 100% coverage and 100% identity with at least 7 adjacent residues in any human protein sequence from the Reference Protein database (refseq) in NCBI. Models, non-redundant RefSeq proteins, and uncultured/environmental sample sequences were excluded in the search to avoid unrelated sequences. The ToxinPred3.0 server (https://webs.iiitd.edu.in/raghava/toxinpred3/) was used to assess potentially toxic peptides. Default threshold score 0.38 was applied in the Hybrid (ET+MERCI) method. This technique demonstrated superiority, achieving 0.98 area under the ROC curve (AUROC) and Matthews correlation coefficient (MCC) 0.81 (Rathore *et al*., 2024). Potential allergens were also identified and excluded using two highly accurate methods: AllerCatPro 2.0 (https://allercatpro.bii.a-star.edu.sg/) and pLM4Alg (https://f6wxpfd3sh.us-east-1.awsapprunner.com). On validation dataset, AllerCatPro2.0 outperformed other benchmark methods for predicting allergenic proteins with maximum performance of area under the curve (AUC) 0.98 with MCC 0.85 on the validation dataset (Sharma *et al.,* 2021). The pLM4Alg tool, another state-of-the-art model, achieved accuracy of 93.4 to 95.1%, MCC 0.869 to 0.902, and AUC scoring of 0.981 to 0.990 (Du *et al.,* 2024). Epitopes identified as non-allergens in both tools were included in the list of candidate epitopes.

Estimation of the population coverage of candidate T-cell epitopes

The cumulative and the individual coverages of candidate CD4+ and CD8+ T-cell epitopes were estimated in world and in regions highly burdened by Mpox, including: North, West, East, South, and Central Africa. The population coverage was also determined for the Southeast Asian (SEA) region. Cumulative population coverage of selected and adjoined CD4+ and CD8+ T-cell epitopes encoded in the mRNA constructs were also calculated to estimate the percentage that can be covered by each mRNA vaccine construct.

Molecular docking of the selected candidate T-cell epitopes

For each Mpox antigen, the epitope with the highest population coverage was selected to be encoded in the mRNA constructs. Using IC_{50} as a metric of binding affinity, the weakest HLA binder was chosen for molecular docking with its respective epitope. The PDB structures of the following HLA molecules were obtained from the RCSB Protein Data Bank (https://www.rcsb.org/). These include: HLA-DRB11501, HLA-DRB10401, HLA-B3501, HLA-B5801, HLA-B5701, HLA-B4001, HLA-A2402, HLA-A0301, HLA-A1101, HLA-B5301, and HLA-B1501, with PDB ID 1bx2, 5ni9, 4lnr, 5vwh, 6bxq, 6iex, 6xqa, 7l1b, 7m8t, 7r7v, and 8elh, respectively. Unrelated hetero atoms were removed from the PDB structures of the HLAs. The selected candidate CD4+ and CD8+ T-cell epitopes were docked with the predicted HLA binders using the GalaxyPepDock server (https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=PEPDOCK). Resulting PDB structures of complex with estimated accuracy below 1.0 were further refined in the GalaxyRefineComplex tool (https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=COMPLEX). The dissociation constant (K_D) and the Gibbs free energy change (ΔG_{bind}) of complex formation at 37^0C were also calculated in the Prodigy webserver (https://rascar.science.uu.nl/prodigy/).

Design and evaluation of the mTcPox RNA sequence

The selected candidate CD8+ epitopes were adjoined to form the mTcPox construct wherein the first residue in the first epitope of the construct begins with methionine. Subsequent epitopes were arranged and adjoined in order to maximize the immunoproteosome cleavage scores; thereby, increasing the chance for the desired CD8+ epitope sequences to be produced upon intracellular immunoproteosome cleavage. The Proteasome Cleavage Prediction Server (iPCPS) with a threshold score of 0.5 was used for this purpose (http://imed.med.ucm.es/Tools/pcps/index.html). The relevant physicochemical properties of the product of mTcPox open reading frame (ORF) were determined using the Protparam webserver (https://web.expasy.org/protparam/). The ORF sequence of mTcPox was reverse translated to RNA sequence in the EMBOSS Backtranseq tool and optimized for translation in *Homo sapiens* (https://www.ebi.ac.uk/jdispatcher/st/emboss_backtranseq). Essential sequences were added to finally form the RNA sequence of mTcPox construct. This study suggests the use of the conventional 5' cap 1 (m7GpppN1m) in the 5' end of the 5' untranslated region (5' UTR). The leader sequence (GAAAAAAAAAAAAGGC) of the 5' UTR is consist of poly A (12 nt), Kozak consensus sequence and the human beta globin sequence (ACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACC), in order to

increase translation efficiency and prevent secondary structure formation at the 5' end (Dhungel *et al,* 2017). Sequence of 5' UTR human beta-globin (1-53) was retrieved from NCBI (NM_000518.5). The reverse translated ORF sequence of mTcPox was added after the 5' UTR region. The ORF sequence is followed by a 3' UTR sequence, which is consist of a stop codon (TAA) and the *Homo sapiens* beta-globin 3' UTR sequence (GCTCGCTTTCTTGCTGTCCAATTTCTATTAAAGGTTCCTTTGTTCCCTAAGTCCA ACTACTAAACTGGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTAAT AAAAAACATTTATTTTCATTGCAA). The beta-globin sequence (492 to 628) of the 3' UTR was also obtained from NCBI (NM_000518.5). Lastly, an ideal length (100 nt) of poly (A) tail was added for synthesizing therapeutic mRNA. This study also suggests the replacement of uridine by pseudouridine (N1-methylpseudouridine) to reduce immune reaction toward mRNA (Ho *et al*., 2024). It is also essential to increase the GC content of the mRNA construct for greater stability and better protein translation efficiency. The translation efficiency of ORF sequence can be estimated using Codon Adaptation Index (CAI). The CAI and the GC content of the optimized ORF were calculated using the VectorBuilder Codon Optimization tool for expression in humans (https://en.vectorbuilder.com/tool/codon-optimization.html). RNA Folding Form v2.3 (https://www.unafold.org/mfold/applications/rna-folding-form-v2.php) was used to calculate and predict the minimum free energy for the optimal secondary structure of mTcPox RNA sequence at 37 degrees Celsius. The GU pairs and isolated bases were avoided at the end of the helices.

Design and evaluation of the mBThPox RNA sequence

The mRNA construct of MBThPox, containing the adjoined B-cell and helper T-cell epitopes, was designed to be translated and secreted for the binding of B-cell epitopes to B-cell receptors (BCRs) and the processing of CD4+ epitopes through the exogenous pathway. The *Homo sapiens* renin signal sequence (MDGWRRMPRWGLLLLLWGSCTFG) was used as the first stretch of amino acid sequence in the ORF of the construct. Selected candidate CD4+ epitopes were added next to the renin signal sequence. The helper T-cell epitopes were arranged in order to maximize the probability of cleavage by proteases known to generate and process CD4+ epitopes (cathepsins D, L, and G). The cleavage score of C-terminus residue of each epitope was determined using the ProsperousPlus tool (http://prosperousplus.unimelb-biotools.cloud.edu.au/index.php/prediction). B-cell peptides were added in the last part of the construct. These peptides cover the overlapping candidate linear B-cell epitopes mapped in this study. The peptides were linked using KK and AAY

linkers to enhance the likelihood of recognition of the candidate B-cell epitopes within the model protein structure of the construct. The signal cleavage site of mBThPox ORF sequence was further assessed through the SignalIP 6.0 server (https://services.healthtech.dtu.dk/services/SignalP-6.0/), to ensure that the renin signal sequence in the construct is recognized and cleaved between positions 23 and 24 by signal peptidase I. The probability that the antigen protein will not be GPI-anchored was also predicted in the NetGPI 1.1 webserver (https://services.healthtech.dtu.dk/services/NetGPI-1.1/). These steps are crucial for the translated product of the mBThPox vaccine to be secreted and not to be anchored within the cell membrane.

The adjoined linear B-cell epitopes in the translated protein structure of mBThPox must be located in disordered and paratope-accessible regions for effective BCR binding and recognition. This study determined the secondary structure composition of the vaccine in order to identify the positions of random coil regions using the GOR4 webserver tool (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html). The protein structure model of mBThPox was generated using the GalaxyTBM tool (https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=TBM) in order to assess the accessibility of adjoined linear B-cell epitope. The model structure was further refined in the GalaxyRefine2 webserver (https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE2), and the generated structure model was inspected through iCn3D tool (https://www.ncbi.nlm.nih.gov/Structure/icn3d/). To validate its quality, the percentage of residues lying within the favored and the disallowed regions were calculated using the Wenglab Ramachandran Plot sever (https://bu.wenglab.org/rama/). The quality factor for non-bonded atomic interactions (ERRAT score) was determined in the UCLA SAVESv.6.0 server (https://saves.mbi.ucla.edu/). The accessibility of adjoined B-cell epitopes in the structure of the translated product of mBThPox was investigated in Ellipro (http://tools.iedb.org/ellipro/). The physicochemical properties of the translated product were also determined using the Protparam tool (https://web.expasy.org/protparam/).

The ORF sequence of mBThPox was reverse translated and optimized in the EMBOSS Backtranseq tool. Similar to the steps conducted to generate the final mRNA sequence of mTcPox, essential elements for mRNA translation were also added to the reverse-translated ORF sequence of mBThPox. The percentage GC content and translation efficiency of its ORF sequence was determined, and the minimum free energy of the optimal secondary structure of mBThPox nucleotide sequence was also calculated.

Results

The highly conserved antigens of Mpox virus

The validity for the use of Shannon variability index requires at least 100 protein sequences per antigen. The number of sequences retrieved from NCBI for A29L, A35R, A42R, B6R, E8L, F13L, H3L, L1R, and M1R are 564, 558, 564, 562, 568, 567, 571, 562 and 562, respectively. The positions of variable residues in Mpox antigens are listed in Supplementary File 1.

The candidate CD8+ and CD4+ T-cell epitopes of Mpox antigens

This study predicted and filtered CD8+ and CD4+ T-cell epitopes according to the set of criteria described in the methodology. Processing with MHC I-IEDB Recommend NetMHCIPan and MHC II-IEDB Recommend NetMHCIIPan 4.1 BA resulted in T-cell epitopes with IC_{50} values ranging from 4.6 to 499.99 and 2.9 to 499.1 nM, for CD4+ and CD8+ epitopes, respectively. Five experimentally validated CD8+ T-cell epitopes were obtained and utilized to provide threshold scores with the aid of three state-of-the-art immunogenicity tools. Table 1 indicates the threshold scores employed in filtering CD8+ T-cell epitopes which are 0.566, -0.104, and 0.043, using DIm, CIIm, and PRIME, respectively. Predicted cytotoxic T-cell epitopes with scores less than any of the 3 thresholds were excluded. Mapped CD4+ T-cell epitopes classified as non-IL4 and/or non-IFN-gamma inducers were also excluded in this study.

Table 1. Immunogenicity scores of experimentally validated cytotoxic T-cell epitopes

CD8+ T-cell epitope	HLA I allele	DIm score	CIIm score	PRIME score
LLMGTLGIV	A*0201	0.820	0.386	0.043
SLYNTVATL	A*0201	0.925	0.131	0.267
YLQPRTFLL	A*0201	0.811	0.125	0.187
TIADYNYKL	A*0201	0.566	0.111	0.212
YIWLGFIAGL	A*0201	0.759	-0.104	0.612
Immunogenicity thresholds		*0.566*	*-0.104*	*0.043*

Upon manual and thorough inspection, some of the predicted epitopes showed significant hits with particular human proteome sequences. These include: spermatogenesis-associated protein 32 (A29L); inaD-like protein isoform 2 (A35R); breast cancer type 2 susceptibility protein isoform 1, and sister chromatid cohesion protein PDS5 homolog A isoform 1 (B6R); cilia- and flagella-associated protein 43, and von Willebrand factor D and EGF domain-containing protein isoform 1 precursor (E8L); E3 SUMO-protein ligase EGR2 isoform a, early growth response protein 3 isoform 1, neuron navigator 3 (F13L); S phase cyclin A-associated protein in the endoplasmic reticulum isoform a, dimethyladenosine transferase isoform 1, and rho guanine nucleotide exchange factor 26 isoform 1 (H3L); and apoptosis-inducing factor short isoform 3, mucin-5B precursor, and E3 ubiquitin-protein ligase (M1R) with the predicted CD4+ T-cell epitopes. One CD8+ B6R epitope was found to have a significant hit with human GPI mannosyltransferase 4. After the exclusion of potentially cross-reactive, allergenic and toxic peptides, a total of 559 CD4+ and 104 CD8+ epitopes remained; and hereby, classified as candidate T-cell epitopes. The IC_{50} value range for the set of candidate CD4+ and CD8+ T-cell epitopes, and the number of candidate epitopes for each Mpox antigen, are also found in Supplementary File 2.

The candidate B-cell epitopes of Mpox EEV proteins

The predicted external regions of Mpox EEV proteins A35R and B6R are positions 58 to 181, and 20 to 227, respectively. Using only the external sequences, consensus B-cell epitopes were determined by utilizing 7 different B-cell epitope mapping tools. Epitopes containing variable and PTM residues were excluded. After filtering out potentially cross-reactive, allergenic, and toxic epitopes, a total 18 (12 for A35R and 6 for B6R) epitopes remained in the list of candidate Mpox B-cell epitopes. Table 2 shows three peptide sequences (P1 to P3) spanning the 18 candidate B-cell epitopes. These three peptides were included in the mRNA vaccine construct proposed by the study. P1 and P2 are found in positions 146-178 and 115-143 of Mpox A35R, respectively. P3 is in position 23-45 of Mpox B6R.

Table 2. The peptide sequences of P1, P2, and P3 with the candidate B-cell epitopes

Antigen	Peptide	Code	Position	Residues	B-cell epitopes overlap
A35R	DYVEDTWGSDGNPITKTTSDYQDSDVSQEVRKY	P1	146-178	33	TKTTSDYQDSDVSQEV DTWGSDGNPITKTTSDYQDSDVSQEVRKYFC TWGSDGNPITKTTSDYQDSDVSQEV ITKTTSDYQDSDV EDTWGSDGNPITKTTSDYQDSDVS DYVEDTWGSDGNPITKTTSDYQDSDVSQEVRKY EDTWGSDGNPITKTTS
	DYKSFEDAKANCAAESSTLPNKSDVLTTW	P2	115-143	29	STLPNKSDVLTT AESSTLPNKSDVLTTW STLPNKSD ESSTLPNKSDVL DYKSFEDAKANCAAESSTLPNKSDV
B6R	VPTMNNAKLTSTETSFNDKQKVT	P3	23-45	23	PTMNNAKLTSTETSFN VPTMNNAKLTSTETSFNDKQKV PTMNNAKLTSTETSFNDKQKV AKLTSTETSFNDKQKV KLTSTETSFNDKQKVT LTSTETSFNDK

Candidate cytotoxic and helper T-cell epitopes cover significant percentage of populations

The candidate epitopes encoded by the couplet mRNA constructs were selected according to the highest percentage of population it can cover. Table 3 presents the selected candidate CD4+ and CD8+ T-cell epitopes, with the IC_{50} range values for their HLA binders. The estimated world population coverage of the set of candidate CD8+ and CD4+ epitopes are 77.67% and 81.81%, respectively. Consequently, the cumulative population coverage of epitopes in the mTcPox construct is 77.67% and 81.81% for the mBThPox mRNA construct (Table 4), as epitopes with the highest population coverage per antigen were selected for inclusion. The individual population coverage of each T-cell epitope in selected regions, where Mpox is currently prevalent, are also indicated in Table 4. The data for some HLA II alleles were not available and therefore not included in the calculation. These alleles include:

HLA-DPA1*03:01/DPB1*04:02, HLA-DPA1*02:01/DPB1*05:01, HLA-DQA1*04:01/DQB1*04:02, HLA-DQA1*05:01/DQB1*02:01, HLA-DRB4*01:01, HLA-DPA1*02:01/DPB1*01:01, HLA-DQA1*01:02/DQB1*06:02, HLA-DRB3*02:02, HLA-DPA1*01:03/DPB1*04:01, HLA-DPA1*02:01/DPB1*14:01, HLA-DPA1*01:03/DPB1*02:01, HLA-DQA1*05:01/DQB1*03:01, and HLA-DRB5*01:01.

The set of CD8+ epitopes can cover as high as 92.77% of the population in West Africa, while the set of CD4+ epitopes can cover as high as 81.81% of the world population. The lowest percentage of populations covered by CD8+ and CD4+ epitope sets are 77.67% and 32.1%, in the world and in South Africa, respectively.

Table 3. The peptide sequence and the IC_{50} range values of selected T-cell epitopes

Antigen	T-cell	Epitope	IC_{50} range (nM)
A29L		LAIPATEFF	14.5—139.6
A35R		TSVFSATVY	136.1—155.6
A42R		IARTHTALIF	78.3—175.1
B6R		MTINCDVGY	27—378.4
E8L	CD8+	MSAPFDSVFY	21.6—336.8
F13L		YQNHGFVSF	6.0—406.9
H3L		FTTPLISFF	42.8—478.4
L1R		QYLLTMFF	78.3—252.8
M1R		ILAALFMYY	33.5—490.9
A29L		EVLFRLENHAETLRA	63.87—329.5
A35R		KRVIGLCIRISMVIS	61.2—412.01
A42R		LKPLIGQKFCIVYTNSL	24.83—395.86
B6R		GGVIHLSCKSGFTLT	173.03—394.33
E8L	CD4+	PSTLDYFTYLGTTINHS	37.74—412.79
F13L		FKAFNSAKNSWLNLC	13.58—184.32
H3L		YPGVMYTFTTPLISF	11.45—85.41
L1R		SLMRFKKESALATTAID	10.33—420.48
M1R		MFTAALNIQTSVNTV	15.59—412.28

Table 4. Population coverage of individual T-cell epitopes encoded in the mRNA vaccine constructs

Epitope	World	East Africa	West Africa	North Africa	South Africa	Central Africa	Southeast Asia
CD4+ A29L	63.74%	50.30%	54.75%	59.37%	7.65%	47.53%	47.99%
CD4+ A35R	66.41%	57.36%	56.40%	60.49%	7.65%	52.75%	50.93%
CD4+ A42R	62.96%	31.28%	46.50%	53.28%	7.65%	38.22%	40.23%
CD4+ B6R	60.37%	53.35%	48.59%	51.52%	32.10%	46.19%	42.57%
CD4+ E8L	66.73%	43.60%	49.05%	53.33%	1.79%	34.42%	46.96%
CD4+ F13L	70.55%	53.08%	57.74%	60.85%	7.65%	49.06%	48.55%
CD4+ H3L	72.95%	59.93%	59.33%	61.95%	7.65%	54.20%	51.48%
CD4+ L1R	66.73%	43.60%	49.05%	53.33%	1.79%	34.42%	46.96%
CD4+ M1R	70.55%	53.08%	57.74%	60.85%	7.65%	49.06%	48.55%
CD4+ set in mBThPox	***81.81%***	***68.30%***	***65.23%***	***75.06%***	***32.10%***	***62.71%***	***56.98%***
CD8+ A29L	29.38%	41.21%	59.50%	41.33%	31.93%	42.87%	21.03%
CD8+ A35R	18.52%	18.92%	31.44%	18.28%	19.24%	22.19%	6.90%
CD8+ A42R	11.72%	12.32%	11.30%	6.42%	6.09%	8.36%	17.35%
CD8+ B6R	42.23%	33.06%	48.97%	33.93%	39.75%	33.14%	52.67%
CD8+ E8L	37.74%	35.28%	45.32%	32.68%	34.99%	37.00%	21.76%
CD8+ F13L	32.60%	23.10%	47.63%	35.34%	24.35%	30.06%	35.55%
CD8+ H3L	32.88%	42.67%	50.89%	35.57%	32.99%	37.61%	30.33%
CD8+ L1R	26.18%	17.08%	31.79%	26.56%	25.59%	19.95%	40.17%
CD8+ M1R	43.99%	26.95%	40.41%	29.53%	35.32%	33.61%	41.44%
CD8+ set in mTcPox	***77.67%***	***68.9%***	***82.72%***	***72.70%***	***76.02%***	***67.23%***	***85.69%***

The T-cell epitopes encoded by the couplet mRNA constructs can potentially dock within the peptide-binding groove of HLA

The epitopes encoded in the mRNA construct were docked to the PDB structures of their weakest (highest IC_{50} value) HLA allele binders. This step was conducted to support the binding data obtained from T-cell epitope prediction tools employed in this research. As depicted in Figures 1 and 2, each candidate epitope docks within the peptide-binding groove of the HLA I and HLA II molecules, respectively.

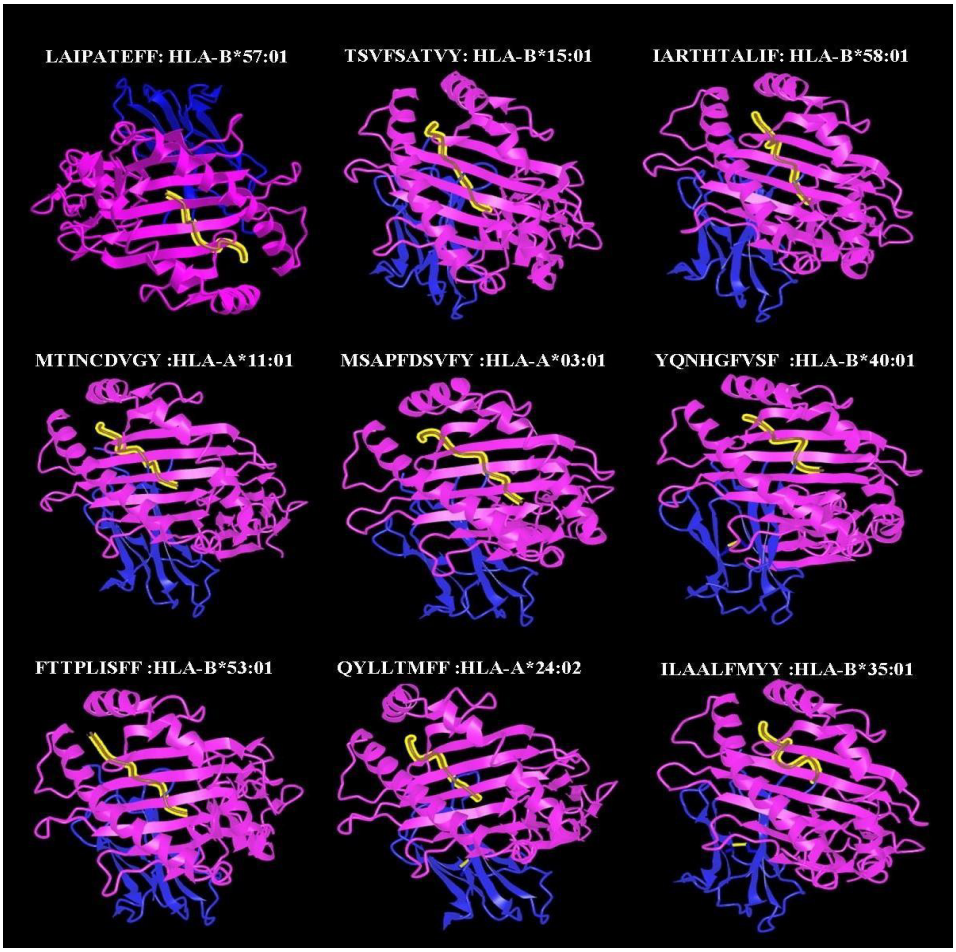


Figure 1. The structure models of CD8+ epitopes docked with the predicted HLA I allele binders. The HLA chains are depicted in blue and magenta while the epitopes docked within the peptide-binding groove are depicted in yellow.

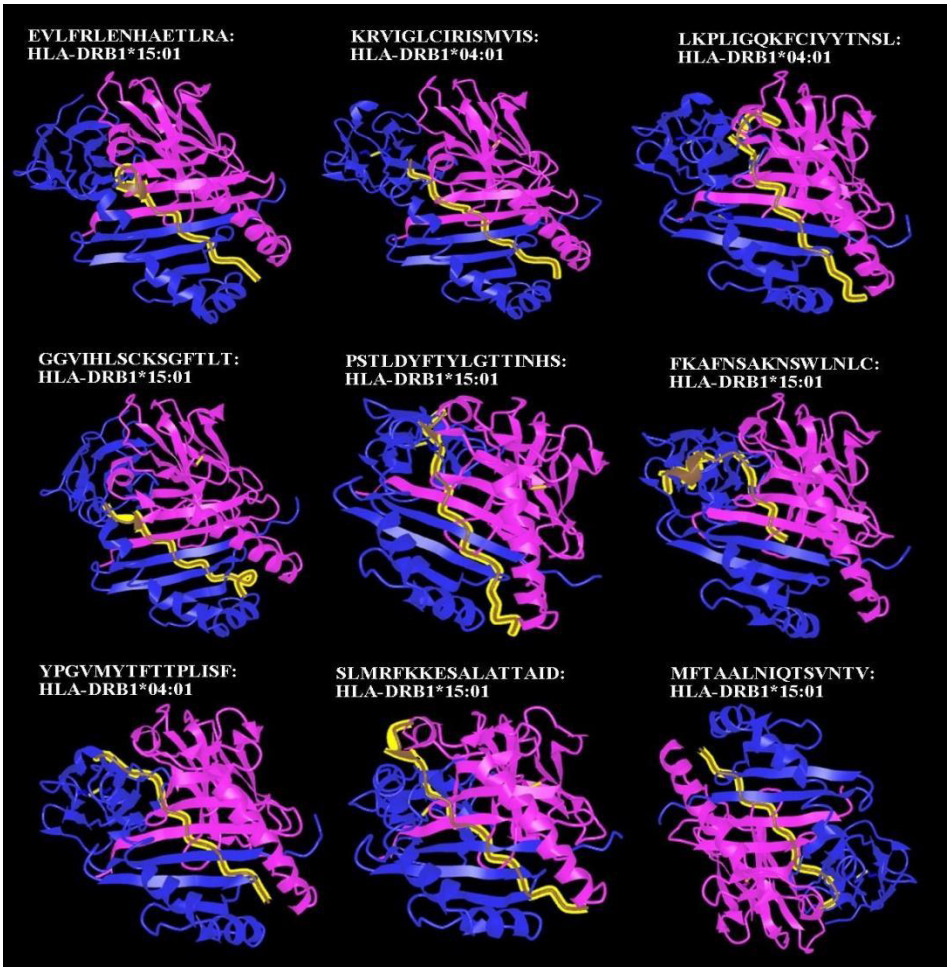


Figure 2. The structure models of CD4+ epitopes docked with the predicted HLA II allele binders. The HLA chains are depicted in blue and magenta while the epitopes docked within the peptide-binding groove are depicted in yellow.

The estimated accuracy (EA) of the initial PDB structures were tabulated in Table 5. For each docked complex, generated PDB structures having an EA of < 1.0 were further refined, while those with accuracy of 1.0 were directly utilized. The percentage of Rama favored residues for the refined structures (RR) were higher than the initial (unrefined) PDB structures (RI), supporting the necessity and the efficacy of refinement steps conducted in this study for structures with EA of < 1.0. The dissociation constant (K_D) and the Gibbs free energy of complex formation (ΔG_{bind}) at 37 ^{0}C, are also indicated in Table 5.

Table 5. The summary of modeling accuracy and binding properties of epitope-HLA structure complexes

Epitope	Weakest HLA binder	IC_{50}	PDB ID	EA	RI	RR	ΔG_{Bind}	K_D
LAIPATEFF	B*57:01	139.6	6BXQ	1	NA	NA	-8.8	6.70E-07
TSVFSATVY	B*15:01	155.6	8ELH	1	NA	NA	-12.5	1.50E-09
IARTHTALIF	B*58:01	175.1	5VWH	1	NA	NA	-10	8.50E-08
MTINCDVGY	A*11:01	378.4	7M8T	1	NA	NA	-9.6	1.60E-07
MSAPFDSVFY	A*03:01	336.8	7L1B	1	NA	NA	-11.4	8.70E-09
YQNHGFVSF	B*40:01	406.9	6IEX	0.981	98.1	98.1	-9.7	1.40E-07
FTTPLISFF	B*53:01	478.4	7R7V	1	NA	NA	-11.2	1.30E-08
QYLLTMFF	A*24:02	252.8	6XQA	0.978	98.7	98.7	-10.3	5.60E-08
ILAALFMYY	B*35:01	490.9	4LNR	1	NA	NA	-11.5	7.30E-09
EVLFRLENHAETLRA	DRB1*15:01	329.5	IBX2	0.846	97.4	99.2	-12.1	3.00E-09
KRVIGLCIRISMVIS	DRB1*04:01	412.01	5NI9	0.903	98.4	99	-11.8	4.50E-09
LKPLIGQKFCIVYTNSL	DRB1*04:01	395.86	5NI9	0.883	98.4	98.7	-13.3	4.10E-10
GGVIHLSCKSGFTLT	DRB1*15:01	394.33	IBX2	0.887	98.9	99.4	-12.4	1.90E-09
PSTLDYFTYLGTTINHS	DRB1*15:01	412.79	IBX2	0.944	98.2	98.4	-13.4	3.70E-10
FKAFNSAKNSWLNLC	DRB1*15:01	184.32	IBX2	0.888	98.4	98.4	-11.2	1.20E-08
YPGVMYTFTTPLISF	DRB1*04:01	85.41	5NI9	0.954	97.6	99	-10.7	3.00E-08
SLMRFKKESALATTAID	DRB1*15:01	420.48	IBX2	0.85	98.7	98.7	-12.1	3.10E-09
MFTAALNIQTSVNTV	DRB1*15:01	412.28	IBX2	0.848	97.9	98.4	-12.6	1.30E-09

***Footnote**: EA is the estimated accuracy. RI (8th column) is the percentage of Rama favored residues in the initial structure and RR (9th column) is the percentage of Rama favored residues in the initial structure. IC_{50}, K_D, and ΔG_{bind}, are expressed in nM, M, and kcal/mol, respectively.*

The couplet mRNA constructs contain adjoined elements and epitopes directed to endogenous and exogenous pathways

Nine selected candidate CD8+ T-cell epitopes were arranged in order to maximize the cleavage scores of epitopes. Sequence of the protein product of mTcPox contains 82 residues (MTINCDVGYFTTPLISFFMSAPFDSVFYIAR THTALIFQYLLTMFFTSVFSATVYILAALFMYYLAIPATEFFYQNHGFVSF). It has a molecular weight of 9496.09 Da, pI of 5.20, aliphatic index of 89.27, and grand average of hydropathicity (GRAVY) index of 0.940. Its extinction coefficient (M^{-1} cm^{-1}) in water at 280 nm is 10430 assuming all pairs of Cys residues form

cystines, and 10430 assuming all Cys residues are reduced. Its estimated half-life is 30 hours in mammalian reticulocytes, in vitro. The instability index is computed to be 38.3 which classified the protein as stable. The cleavage scores of adjoined cytotoxic T-cell epitopes are shown in Table 6.

Table 6. The immunoproteosome cleavage scores of adjoined CD8+ epitopes in the mTcPox construct

Adjacent CD8+ T-cell epitopes	Immunoproteosome score
MTINCDVGY—FTTPLISFF	0.5310
FTTPLISFF—MSAPFDSVFY	0.5199
MSAPFDSVFY—IARTHTALIF	0.5557
IARTHTALIF—QYLLTMFF	0.6519
QYLLTMFF—TSVFSATVY	0.8407
TSVFSATVY—ILAALFMYY	0.5557
ILAALFMYY—LAIPATEFF	0.6021
LAIPATEFF—YQNHGFVSF	0.4258

The protein product of mBThPox construct contains 257 residues and was designed to express the human renin signal sequence. Figure 3 shows the sequence of the product of the construct. The amino acid residues highlighted in yellow, gray, blue, green, and orange, are the human renin signal sequence, the series of adjoined CD4+ T-cell epitopes, the B-cell peptide (P1) spanning residues 146-178 of A35R, the B-cell peptide (P2) spanning 115th to 143rd residue of A35R, and the B-cell peptide (P3) covering residues 23-45 of the B6R, respectively. Letters in lower case represent the linkers used in concatenating the three B-cell peptides.

Figure 3. Sequence of the protein product of mBThPox construct.

The predicted signal cleavage site for SecSPI is between positions 23 and 24, with the probability of 0.943327 and likelihood of 0.9988. This cleavage site is located exactly between the last residue of the human renin signal sequence and the first residue of the first CD4+ epitope included in the construct (PSTLDYFTYLGTTINHS). The likelihood that it is not GPI-anchored is 0.995. The next crucial step is to determine whether the designed CD4+ epitopes can be recognized and cleaved by known epitope-generating proteases (cathepsins D, L, and G) of the exogenous pathway. The candidate helper T-cell epitopes were arranged to maximize cathepsin D, L and G cleavage scores of C-terminus residue of each epitope (Table 7).

Table 7. The cathepsin cleavage scores of adjoined CD4+ T-cell epitopes in mBThPox construct

Adjoined CD4+ epitopes	Cleavage scores	Cathepsin
PSTLDYFTYLGTTINHS/GGVIHLSCKSGFTLT	0.212	G
GGVIHLSCKSGFTLT/KRVIGLCIRISMVIS	0.706	L
KRVIGLCIRISMVIS/YPGVMYTFTTPLISF	0.616	L
YPGVMYTFTTPLISF/MFTAALNIQTSVNTV	0.892	D
	0.873	G
MFTAALNIQTSVNTV/SLMRFKKESALATTAID	0.377	G
SLMRFKKESALATTAID/LKPLIGQKFCIVYTNSL	0.723	L
LKPLIGQKFCIVYTNSL/EVLFRLENHAETLRA	0.581	D
EVLFRLENHAETLRA/FKAFNSAKNSWLNLC	0.616	D
FKAFNSAKNSWLNLC/kkDYVEDTWGSDGNPITKTTS DYQDSDVSQEVRKY (KK linker and P1)	0.621	L

The protein structure model of its product was generated and utilized to assess the accessibility of adjoined B-cell epitopes in the mBThPox construct. The three peptide sequences (P1, P2, and P3) cover all 18 candidate B-cell epitopes identified in this study. Two peptides from A35R and 1 peptide from

B6R sequence were arranged and linked using KK and AAY linkers to maximize the accessibility of B-cell epitopes in the structure. The Ramachandran plot analysis of the model structure generated in this study, indicates that the percentage of residues in the highly preferred conformation increased from 93.1% to 94.7%, after structural refinement. Moreover, the clash score of the model structure was reduced from 17.0 to 12.2. The overall quality (ERRAT score) of the refined structure is 88.4298.

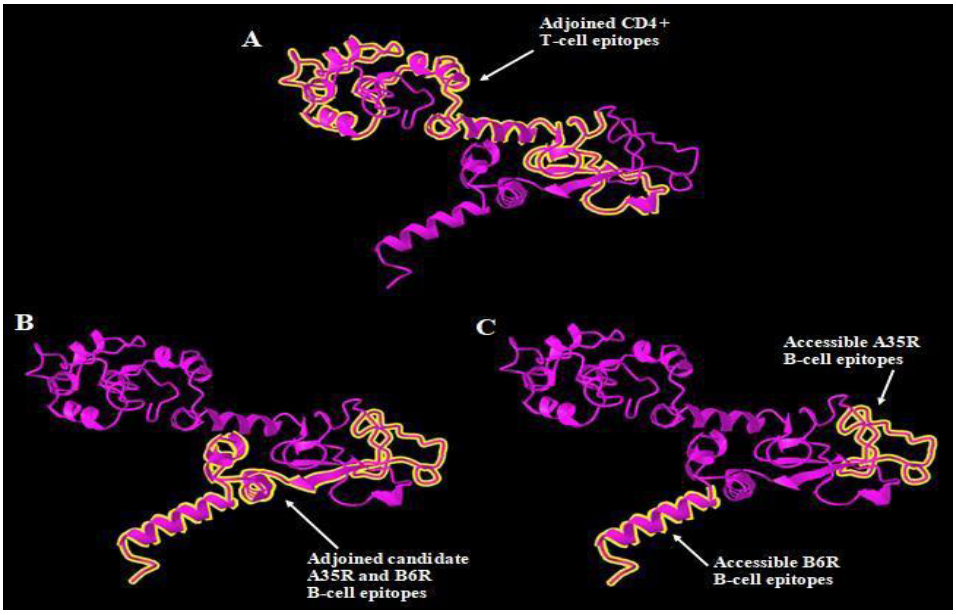


Figure 4. Structure model generated for the protein product of mBThPox construct. Topographical positions of adjoined CD4+ T-cell epitopes (A), adjoined candidate B-cell epitopes (B), and accessible A35R and B6R B-cell epitopes (C) in the model structure are highlighted in yellow.

Structural B-cell epitope analysis in Ellipro predicted the following accessible positions in the generated protein model: 162-198 and 23-257 as linear B-cell epitopes while residues 162-197, 218-225, and 237-257 as structural B-cell epitopes. Out of 18 candidate B-cell epitopes mapped in the study, 8 candidate B-cell epitopes (Table 8) were predicted to be accessible in the protein structure of the vaccine, all of which are accessible both as linear and structural B-cell epitopes. Five and three B-cell epitopes are from A35R and B6R Mpox EEV proteins, respectively. Figure 4 depicts the positions of epitopes in the model

structure generated for the protein product of mBThPox. Moreover, results from the secondary structure analysis imply that the product is made up primarily of random coils (46.69%), and the series of adjoined candidate B-cell epitopes (167 to 257) is located in regions rich in random coils and extended strands.

Table 8. Accessible candidate B-cell epitopes encoded in mBThPox construct

Antigen	Accessible sequences in the model structure	Position in the construct	Accessible candidate B-cell epitopes
A35R	DYVEDTWGSDGNPITKT TSDYQDSDVSQEVRK	162-198	TKTTSDYQDSDVSQEV ITKTTSDYQDSDV TWGSDGNPITKTTSDYQDSDVSQEV EDTWGSDGNPITKTTSDYQDSDVS EDTWGSDGNPITKTTS
B6R	TMNNAKLTSTETSFND KQKVT	237-257	AKLTSTETSFNDKQKV KLTSTETSFNDKQKVT LTSTETSFNDK

Evaluation of the physicochemical properties of the product of mBThPox showed that the protein has a pI of 8.80, molecular weight of 28795.96 Da and aliphatic index of 78.95. Its grand average of hydropathicity (GRAVY) score (-0.140) indicates that the product has more tendency to interact in a hydrophilic environment. The extinction coefficients (M^{-1} cm^{-1}) measured in water at 280 nm are 49765, assuming all pairs of Cys residues form cystines; and 49390, assuming all Cys residues are reduced. Its estimated half-life in mammalian reticulocytes in vitro is 30 hours. The instability index of the product is 31.96, which classified the protein as stable.

The optimized sequences of Mpox couplet mRNA vaccine

The optimized mRNA sequences of mTcPox (A) and mBThPox (B) constructs are shown in Figure 5. The RNA sequence of mTcPox and mBThPox, has a GC content of 57.83% with CAI of 0.99 and a GC content of 61.76% with CAI of 1.0 in humans, respectively. The estimated minimum free energy change of RNA folding for mTcPox and mBThPox at 37 ^{0}C, are -90.10 kcal/mol and -292.90 kcal/mol, respectively.

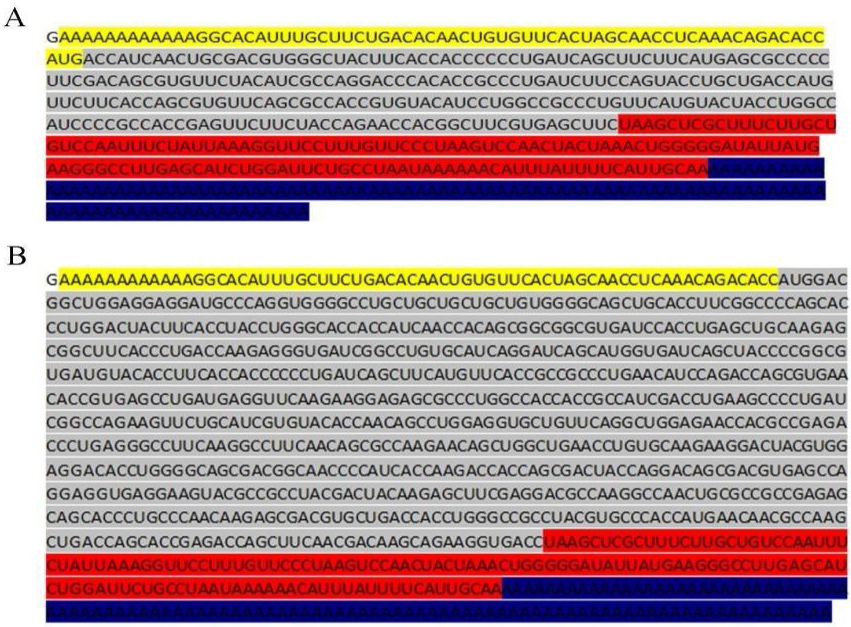


Figure 5. mRNA sequences of the Mpox couplet vaccine. The RNA sequence of mTcPox (A) contains 549 nt while mBThPox (B) has 1074 nt. Both sequences begin with the conventional 5' cap (m7GpppN1m). The 5' UTR sequence of *Homo sapiens* beta-globin which includes the 12 A nucleotides and the Kozak consensus sequence (GGC), is highlighted in yellow. Adjacent to it is the open reading frame (highlighted in gray), followed by the stop codon in the 3' UTR beta-globin sequence (highlighted in red), and a 100-nucleotide poly (A) tail in blue.

Discussion

The declaration of Mpox as a Public Health Emergency of International Concern highlights the urgent need for prevention. Current vaccines use vaccinia, a virus similar to smallpox. JYNNEOS, a third-generation, non-replicating vaccine, shows 85.9% efficacy (95% CL, 73.8–92.4) with fewer adverse reactions (Dalton *et al.,* 2023). ACAM2000, a second-generation vaccine, is replication-competent and linked to more severe side effects. About 95% of recipients develop neutralizing antibodies post-vaccination. LC16, another third-generation vaccine, uses a live attenuated Lister strain; its efficacy is still under study (Wahid *et al*., 2025). Given the limited protection and adverse reactions, Mpox-specific vaccine development is urgently needed.

mRNA vaccine technology offers key advantages over conventional formulations. Notably, mRNA vaccines are safer as they lack live virus and are easily eliminated by the body. They also allow faster development and manufacturing (Koppu *et al*., 2022). This study aimed to design a couplet mRNA vaccine against Mpox disease via immunoinformatics for a streamlined, cost-effective approach. The vaccine includes epitopes that can be potentially recognized by B and T lymphocytes, enabling targeted immune responses. Epitope-based vaccines offer specificity and potency while minimizing biohazard risks (Hajissa *et al*., 2019). Epitope processing varies by type: cytotoxic T-cell epitopes follow the endogenous pathway and are presented via MHC I, while helper T-cell epitopes use the exogenous pathway and are presented by antigen-presenting cells through MHC II. B-cell epitopes bind to B-cell receptors (BCRs), are processed exogenously, and presented to T-cell receptors (TCRs) to activate naive B cells (Blum *et al.,* 2013). The couplet mRNA vaccine includes two constructs: mTcPox (CD8+ T-cell epitopes via endogenous pathway) and mBThPox (B-cell and CD4+ T-cell epitopes via exogenous pathway). This design enhances epitope presentation and immune response. Targeting nine virulent Mpox antigens (A29L, A35R, A42R, B6R, E8L, F13L, H3L, L1R, and M1R) supports a multi-targeted, cost-effective vaccine strategy.

All candidate epitopes obtained in this study possess Shannon variability scores ≤ 0.1, implying high conservancy of the identified epitopes. Its application was validated by utilizing > 500 sequences per Mpox antigen of the same length. These numbers are higher than the required minimum number of sequences (100) to meet the assumption in indicating conservancy using the Shannon variability index (Garcia-Boronat *et al*., 2008). Results from conservancy analysis showed that the full-length sequence of Mpox A42R is highly conserved, while all the other antigens contain at least 1 variable amino acid. F13L has the highest number of variable residues (6) among the Mpox antigens investigated in this study. Nevertheless, these results imply that the sequences of the 9 Mpox antigens are conserved, as there are only few residues with Shannon variability scores > 0.1. Using the highly conserved sequences for epitope mapping is a pivotal step to avoid epitope immune evasion. The highly conserved cytotoxic and helper T-cell epitopes possess acceptable to high binding affinity, as indicated by IC_{50} values ranging from 499.9 to 2.9 nM. Epitopes with IC_{50} < 50 nM and < 500 nM are classified to have high affinity and intermediate binding, respectively (Wu *et al*., 2019).

The ability to induce immunogenicity is an essential property for epitopes included in any vaccine formulation. In this study, the potentially immunogenic cytotoxic T-cell epitopes were identified through the combination of immunogenicity scores generated from 5 experimentally validated CD8+ T-cell epitopes. This

step was conducted due to the current lack of consensus immunogenicity threshold that can be used in immunogenicity tools to distinguish immunogenic from non-immunogenic cytotoxic T-cell epitopes. This strategy may reduce bias by relying on the combination of immunogenicity scores generated from multiple experimentally validated epitopes using multiple state-of-the-art immunogenicity tools, rather than just utilizing a single immunogenicity tool.

Interferon-gamma (IFN-γ) is a key cytokine in promoting the differentiation and function of naive helper T cells (Th0) to Th1 cells, essential for fighting intracellular pathogens, such as Mpox viruses. Interleukin-4 (IL4) is another type of cytokine that induces the activation and differentiation of Th0 to Th2 cells which is crucial in promoting the activation and function of B cells (Aguiar *et al.,* 2024). The candidate CD4+ epitopes generated by this study are classified as IFN-γ and IL4 inducers, implying that the epitopes can potentially induce activation of both Th1 and Th2 cells to fight Mpox-infected cells and activate B cells for antibody production against extracellular Mpox viruses. Overall, 559 candidate CD4+ and 104 candidate CD8+ T-cell epitopes from the nine virulent Mpox antigens were obtained by this study through immunoinformatics approach.

Global concern over Mpox began with its high burden in Africa, marked by rising cases, fatalities, and international spread (Ndembi *et al*., 2025). Although Southeast Asia (SEA) currently has lower prevalence, the emergence of clade Ib poses risks to regions affected by clade IIb (A *et al*., 2025). Strategic development of immunotherapeutics is essential. The WHO recommends 80% population coverage for orthopoxvirus eradication via vaccination (Taube *et al*., 2023). This study estimates that selected CD8+ and CD4+ T-cell epitopes cover 77.67% and 81.81% globally, 92.77% and 75.06% in Africa, and 90.87% and 56.98% in SEA. South Africa shows the lowest CD4+ coverage (32.1%), but CD8+ epitopes compensate with 88.79%. To optimize mRNA vaccine coverage, epitopes with highest global reach were selected: mTcPox covers 81.81%, meeting WHO's target, while mBThPox covers 77.67%, falling short by under 3%. However, this doesn't imply ineffectiveness, as Mpox is less fatal than smallpox and may require lower coverage. The selected epitopes are promiscuous, enabling broader population reach. Actual coverage may exceed estimates, as some HLA alleles were excluded from IEDB Population Coverage tool calculations.

T-cell epitopes encoded in the couplet mRNA constructs were docked to their weakest HLA allele binder. Analysis of the quality of PDB structure models generated for the docked complex showed a high percentage of Rama favored residues and high accuracy. These results supported the use of generated model structures in subsequent evaluation steps. Molecular docking analysis revealed that the selected candidate epitopes bind exactly to the peptide binding groove of the HLA molecule. The favorability of complex formation can be measured

using the equilibrium K_D, wherein smaller K_D indicates favorable binding. In biochemical systems, a complex with K_D value < 1.0e-07 M, is considered to have a high binding affinity (Yewdell, 2022). Docking analysis revealed that the K_D values of complexes formed are all lower than 1.0e-07 M, despite docking with its weakest HLA binder. Moreover, ΔG_{bind} values of the docked complexes range from -8.8 to -13.4 kcal/mol. This is lower than the most commonly used ΔG_{bind} value (-6.0 kcal/mol) for the selection of potential candidates in drug design (Ivanova *et al*., 2022). These results further support the potential immunogenicity of candidate T-cell epitopes mapped in this study.

The Mpox virus has two infectious forms: extracellular enveloped virion (EEV) and intracellular mature virion (IMV). EEV drives transmission and has three surface proteins: A35R (type II membrane protein), B6R (glycoprotein regulating the complement system), and F13L/CL19 (inner surface protein) (Sagdat *et al*., 2024). F13L was excluded from B-cell epitope mapping due to its buried location. A35R and B6R are neutralizing and ideal vaccine targets (Yang *et al.,* 2025). Epitope mapping considered flexibility, accessibility, hydrophilicity, secondary structure, and antigenicity. All candidate epitopes are in conserved, external regions of A35R and B6R, allowing BCR binding. Predicted epitopes with post-translational modifications (PTMs) were excluded, as unmodified vaccine residues may hinder recognition. None of the 18 candidate epitopes contain PTM residues.

Arrangement of nine candidate CD8+ T-cell epitopes in mTcPox construct showed, that each cytotoxic epitope can be cleaved and generated with high probability, as indicated by immunoproteosome cleavage scores > 0.5000, except for the last epitope (YQNHGFVSF) which has an immunoproteosome score of 0.4258. Investigation of its physicochemical properties in silico, suggest that the product is stable in vitro, with relatively longer half-life in mammalian cells, and a preference for interactions with hydrophobic systems.

Results from the *in silico* analysis suggest the mBThPox protein is secreted via the Sec translocon and cleaved by Signal Peptidase I (SecSPI), without membrane anchoring. Its CD4+ epitopes are predicted to be processed by cathepsins D, L, and G in the exogenous pathway. KK and AAY linkers were found to enhance B-cell epitope accessibility and were used to join three peptides covering 18 candidate epitopes. The modeled protein structure is high quality, with >90% residues in favored conformations and an ERRAT score >50% (Gupta *et al*., 2013; Lombard *et al*., 2024). Eight B-cell epitopes (5 from A35R, 3 from B6R) are accessible as linear and structural epitopes. However, detection may be limited by the PDB model, Ellipro tool, and study thresholds. The remaining nine epitopes might be identified using alternative prediction tools with relaxed criteria. Physicochemical analysis indicates the protein is stable with a half-life >24 hours in human reticulocytes.

This study recommends using a conventional 5' cap (m7GpppN1m) with a 7-methylguanosine moiety to protect mRNA from exonuclease degradation and enhance splicing, nuclear export, and translation (Gote *et al*., 2023). The 5' UTR of human beta-globin improves translational efficiency in mRNA vaccines (Ma *et al*., 2024). Incorporating a Kozak sequence and 12 adenosines (polyA) into the leader sequence boosts translation and prevents secondary structure formation at the 5' end, as shown in poxvirus mRNA (Dhungel *et al*., 2017). The 3' UTR of human beta-globin increases protein yield and mRNA stability (Adibzadeh *et al*., 2019), while a 100-nucleotide poly(A) tail further enhances stability (Vélez *et al*., 2025). The couplet mRNA vaccines (mTcPox and mBThPox) show high potential for human expression, with CAI values of 0.99 and 1.00, respectively. Optimal GC content may protect against endoribonuclease degradation and prolong mRNA lifespan (Leong *et al*., 2025). Structural analysis reveals that neither mRNA begins with a hairpin loop, facilitating efficient ribosome binding. Their minimum free energy (MFE) values support stable folding and accessibility for translation.

More than its sophisticated design and promising efficacy, the B-cell and T-cell epitopes that can be expressed and generated from the Mpox couplet mRNA vaccine, exhibited very low probability of causing adverse reactions due to cross-reactivity with human proteomes, allergic reactions, and toxicity. Results from in silico safety profile evaluation suggest its safety for future Mpox vaccine clinical applications.

Conclusions

This study offers a novel and specific couplet mRNA vaccine design for future research and vaccine development against the PHEIC disease-causing Mpox virus. Leveraging computational immunology enables efficient, cost-effective discovery of immunotherapeutics for infectious diseases like Mpox. However, its application demands caution due to some inherent assumptions and restrictions. A limited set of validated cytotoxic epitopes was used to establish thresholds for excluding non-immunogenic candidates, highlighting the need to refine consensus immunogenicity values across tools. Overall, this is the first strategic design of a couplet mRNA vaccine containing potentially immunogenic B-cell and T-cell epitopes against multiple Mpox antigens. Safety assessments *in silico* suggest that it can be administered in humans with very minimal adverse reactions. Further *in vitro*, animal, and clinical studies are anticipated.

Ethical Approval
This study did not conduct procedures involving human participants, samples and data; thus, ethics approval was not deemed necessary.

Declaration of Conflict of Interest
The author declares no conflict of interest.

References

A, J., Gan, G., Endo, A., Jin Tan, R. K., Prem, K., & Dickens, B. L. (2025). Is Southeast Asia and the Western Pacific ready for potential monkeypox virus outbreaks?. *The Lancet regional health. Western Pacific,* 57, 101526. https://doi.org/10.1016/j.lanwpc.2025.101526

Adibzadeh, S., Fardaei, M., Takhshid, M. A., Miri, M. R., Rafiei Dehbidi, G., Farhadi, A., Ranjbaran, R., Alavi, P., Nikouyan, N., Seyyedi, N., Naderi, S., Eskandari, A., & Behzad-Behbahani, A. (2019). Enhancing Stability of Destabilized Green Fluorescent Protein Using Chimeric mRNA Containing Human Beta-Globin 5' and 3' Untranslated Regions. *Avicenna J Med Biotechnol,* 11(1), 112–117. https://pmc.ncbi.nlm.nih.gov/articles/PMC6359690/

Aguiar MPd & Vieira JH. (2024). Entrance to the multifaceted world of CD4+ T cell subsets. *Explor Immunol,* 4, 152–68. https://doi.org/10.37349/ei.2024.00134

Beeson, A., Styczynski, A., Hutson, C. L., Whitehill, F., Angelo, K. M., Minhaj, F. S., Morgan, C., Ciampaglio, K., Reynolds, M. G., McCollum, A. M., & Guagliardo, S. A. J. (2023). Mpox respiratory transmission: the state of the evidence. *Lancet Microbe*, 4(4), e277–e283. https://doi.org/10.1016/S2666-5247(23)00034-4

Blum, J. S., Wearsch, P. A., & Cresswell, P. (2013). Pathways of antigen processing. *Annu Rev Immunol*, 31, 443–473. https://doi.org/10.1146/annurev-immunol-032712-095910

Bogacka, A., Wroczynska, A., Rymer, W., Grzesiowski, P., Kant, R., Grzybek, M., & Parczewski, M. (2025). Mpox unveiled: Global epidemiology, treatment advances, and prevention strategies. *One Health*, 20, 101030. https://doi.org/10.1016/j.onehlt.2025.101030

Callaby, H., Belfield, A., Otter, A. D., Atkinson, B., Reynolds, M., Roberts, H., & Gordon, N. C. (2025). Mpox: current knowledge and understanding-a scoping review. *FEMS Microbiol Rev*, 49, fuaf025. https://doi.org/10.1093/femsre/fuaf025

Dalton, A. F., Diallo, A. O., Chard, A. N., Moulia, D. L., Deputy, N. P., Fothergill, A., Kracalik, I., Wegner, C. W., Markus, T. M., Pathela, P., Still, W. L., Hawkins, S., Mangla, A. T., Ravi, N., Licherdell, E., Britton, A., Lynfield, R., Sutton, M., Hansen, A. P., Betancourt, G. S., ... CDC Multijurisdictional Mpox Case Control Study Group (2023). Estimated Effectiveness of JYNNEOS Vaccine in Preventing Mpox: A Multijurisdictional Case-Control Study - United States, August 19, 2022-March 31, 2023. *MMWR Morb Mortal Wkly Rep*, 72(20), 553–558. https://doi.org/10.15585/mmwr.mm7220a3

Dhanda, S. K., Vir, P., & Raghava, G. P. (2013). Designing of interferon-gamma inducing MHC class-II binders. *Biol Direct*, 8, 30. https://doi.org/10.1186/1745-6150-8-30

Dhungel, P., Cao, S., & Yang, Z. (2017). The 5'-poly(A) leader of poxvirus mRNA confers a translational advantage that can be achieved in cells with impaired cap-dependent translation. *PLoS Pathog*, 13(8), e1006602. https://doi.org/10.1371/journal.ppat.1006602

Du, Z., Xu, Y., Liu, C., & Li, Y. (2024). pLM4Alg: Protein Language Model-Based Predictors for Allergenic Proteins and Peptides. *J. Agric. Food Chem.*, 72(1), 752–760. https://doi.org/10.1021/acs.jafc.3c07143

Garcia-Boronat, M., Diez-Rivero, C. M., Reinherz, E. L., & Reche, P. A. (2008). PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. *Nucleic Acids Res.*, 36(Web Server issue), W35–W41. https://doi.org/10.1093/nar/gkn211

Gfeller, D., Schmidt, J., Croce, G., Guillaume, P., Bobisse, S., Genolet, R., Queiroz, L., Cesbron, J., Racle, J., & Harari, A. (2023). Improved predictions of antigen presentation and TCR recognition with MixMHCpred2.2 and PRIME2.0 reveal potent SARS-CoV-2 CD8+ T-cell epitopes. *Cell Syst.*, 14(1), 72–83.e5. https://doi.org/10.1016/j.cels.2022.12.002

Gote, V., Bolla, P. K., Kommineni, N., Butreddy, A., Nukala, P. K., Palakurthi, S. S., & Khan, W. (2023). A Comprehensive Review of mRNA Vaccines. *Int. J. Mol. Sci.*, 24(3), 2700. https://doi.org/10.3390/ijms24032700

Gupta, C. L., Akhtar, S., Bajpaib, P., Kandpal, K. N., Desai, G. S., & Tiwari, A. K. (2013). Computational modeling and validation studies of 3-D structure of neuraminidase protein of H1N1 influenza A virus and subsequent in silico elucidation of piceid analogues as its potent inhibitors. *EXCLI J.*, 12, 215–225

Hajissa, K., Zakaria, R., Suppian, R., & Mohamed, Z. (2019). Epitope-based vaccine as a universal vaccination strategy against Toxoplasma gondii infection: A mini-review. *J. Adv. Vet. Anim. Res.*, 6(2), 174–182. https://doi.org/10.5455/javar.2019.f329

Ho, L. L. Y., Schiess, G. H. A., Miranda, P., Weber, G., & Astakhova, K. (2024). Pseudouridine and N1-methylpseudouridine as potent nucleotide analogues for RNA therapy and vaccine development. *RSC Chem. Biol.*, 5(5), 418–425. https://doi.org/10.1039/d4cb00022f

Ivanova, L., & Karelson, M. (2022). The Impact of Software Used and the Type of Target Protein on Molecular Docking Accuracy. *Molecules*, 27(24), 9041. https://doi.org/10.3390/molecules27249041

Koppu, V., Poloju, D., Puvvala, B., Madineni, K., Balaji, S., Sheela, C. M. P., Manchikanti, S. S. C., & Moon, S. M. (2022). Current Perspectives and Future Prospects of mRNA Vaccines against Viral Diseases: A Brief Review. *Int J Mol Cell Med*, 11(3), 260–272. https://doi.org/10.22088/IJMCM.BUMS.11.3.260

Leong, K. Y., Tham, S. K., & Poh, C. L. (2025). Revolutionizing immunization: a comprehensive review of mRNA vaccine technology and applications. *Virol J*, 22(1), 71. https://doi.org/10.1186/s12985-025-02645-6

Li, G., Iyer, B., Prasath, V. B. S., Ni, Y., & Salomonis, N. (2021). DeepImmuno: deep learning-empowered prediction and generation of immunogenic peptides for T-cell immunity. *Brief Bioinform*, 22(6), bbab160. https://doi.org/10.1093/bib/bbab160

Lombard, V., Grudinin, S., & Laine, E. (2024). Explaining Conformational Diversity in Protein Families through Molecular Motions. *Sci Data*, 11(1), 752. https://doi.org/10.1038/s41597-024-03524-5

Lu, J., Xing, H., Wang, C., Tang, M., Wu, C., Ye, F., Yin, L., Yang, Y., Tan, W., & Shen, L. (2023). Mpox (formerly monkeypox): pathogenesis, prevention, and treatment. *Signal Transduct Target Ther*, 8(1), 458. https://doi.org/10.1038/s41392-023-01675-2

Ma, Q., Zhang, X., Yang, J., Li, H., Hao, Y., & Feng, X. (2024). Optimization of the 5' untranslated region of mRNA vaccines. *Sci Rep*, 14(1), 19845. https://doi.org/10.1038/s41598-024-70792-x

Ndembi, N., Folayan, M. O., Komakech, A., Mercy, K., Tessema, S., Mbala-Kingebeni, P., Ngandu, C., Ngongo, N., Kaseya, J., & Abdool Karim, S. S. (2025). Evolving Epidemiology of Mpox in Africa in 2024. *N Engl J Med*, 392(7), 666–676. https://doi.org/10.1056/NEJMoa2411368

Nebangwa, D. N., Shey, R. A., Shadrack, D. M., Shintouo, C. M., Yaah, N. E., Yengo, B. N., Efeti, M. T., Gwei, K. Y., Fomekong, D. B. A., Nchanji, G. T., Lemoge, A. A., Ntie-Kang, F., & Ghogomu, S. M. (2024). Predictive immunoinformatics reveal promising safety and anti-onchocerciasis protective immune response profiles to vaccine candidates (Ov-RAL-2 and Ov-103) in anticipation of phase I clinical trials. *PLoS One*, 19(10), e0312315. https://doi.org/10.1371/journal.pone.0312315

Nibeyro, G., Baronetto, V., Folco, J. I., Pastore, P., Girotti, M. R., Prato, L., Morón, G., Luján, H. D., & Fernández, E. A. (2023). Unraveling tumor specific neoantigen immunogenicity prediction: a comprehensive analysis. *Front Immunol*, 14, 1094236. https://doi.org/10.3389/fimmu.2023.1094236

Olotu, F. A., & Soliman, M. E. S. (2021). Immunoinformatics prediction of potential B-cell and T-cell epitopes as effective vaccine candidates for eliciting immunogenic responses against Epstein-Barr virus. *Biomed J*, 44(3), 317–337. https://doi.org/10.1016/j.bj.2020.01.002

Rathore, A. S., Choudhury, S., Arora, A., Tijare, P., & Raghava, G. P. S. (2024). ToxinPred 3.0: An improved method for predicting the toxicity of peptides. *Comput Biol Med*, 179, 108926. https://doi.org/10.1016/j.compbiomed.2024.108926

Rawat, S. S., Keshri, A. K., Kaur, R., & Prasad, A. (2023). Immunoinformatics Approaches for Vaccine Design: A Fast and Secure Strategy for Successful Vaccine Development. *Vaccines*, 11(2), 221. https://doi.org/10.3390/vaccines11020221

Reardon, B., Koşaloğlu-Yalçın, Z., Paul, S., Peters, B., & Sette, A. (2021). Allele-Specific Thresholds of Eluted Ligands for T-Cell Epitope Prediction. *Mol Cell Proteomics*, 20, 100122. https://doi.org/10.1016/j.mcpro.2021.100122

Sagdat, K., Batyrkhan, A., & Kanayeva, D. (2024). Exploring monkeypox virus proteins and rapid detection techniques. *Front Cell Infect Microbiol*, 14, 1414224. https://doi.org/10.3389/fcimb.2024.1414224

Sharma, N., Patiyal, S., Dhall, A., Pande, A., Arora, C., & Raghava, G. P. S. (2021). AlgPred 2.0: an improved method for predicting allergenic proteins and mapping of IgE epitopes. *Brief Bioinform*, 22(4), bbaa294. https://doi.org/10.1093/bib/bbaa294

Taube, J. C., Rest, E. C., Lloyd-Smith, J. O., & Bansal, S. (2023). The global landscape of smallpox vaccination history and implications for current and future orthopoxvirus susceptibility: a modelling study. *Lancet Infect Dis*, 23(4), 454–462. https://doi.org/10.1016/S1473-3099(22)00664-8

Vélez, D. E., Torres, B. L., & Hernández, G. (2025). The Bright Future of mRNA as a Therapeutic Molecule. *Genes*, 16(4), 376. https://doi.org/10.3390/genes16040376

Wahid, M., Mandal, R. K., Sikander, M., Khan, M. R., Haque, S., Nagda, N., Ahmad, F., & Rodriguez-Morales, A. J. (2025). Safety and Efficacy of Repurposed Smallpox Vaccines Against Mpox: A Critical Review of ACAM2000, JYNNEOS, and LC16. *J Epidemiol Glob Health*, 15(1), 88. https://doi.org/10.1007/s44197-025-00432-8

Wu, T., Guan, J., Handel, A., Tscharke, D. C., Sidney, J., Sette, A., Wakim, L. M., Sng, X. Y. X., Thomas, P. G., Croft, N. P., Purcell, A. W., & La Gruta, N. L. (2019). Quantification of epitope abundance reveals the effect of direct and cross-presentation on influenza CTL responses. *Nat Commun*, 10(1), 2846. https://doi.org/10.1038/s41467-019-10661-8

Yadav, R., Chaudhary, A. A., Srivastava, U., Gupta, S., Rustagi, S., Rudayni, H. A., Kashyap, V. K., & Kumar, S. (2025). Mpox 2022 to 2025 Update: A Comprehensive Review on Its Complications, Transmission, Diagnosis, and Treatment. *Viruses*, 17(6), 753. https://doi.org/10.3390/v17060753

Yang, X., Guo, L., Duan, H., Fan, M., Xu, F., Chi, X., Pan, S., Liu, X., Zhang, X., Gao, P., Zhang, F., Wang, X., Guo, F., Ge, J., Ren, L., & Yang, W. (2025). Identification of neutralizing nanobodies protecting against poxvirus infection. *Cell Discov*, 11(1), 31. https://doi.org/10.1038/s41421-025-00771-7

Yewdell J. W. (2022). MHC Class I Immunopeptidome: Past, Present, and Future. *Mol Cell Proteomics*, 21(7), 100230. https://doi.org/10.1016/j.mcpro.2022.100230

Zahmatyar, M., Fazlollahi, A., Motamedi, A., Zolfi, M., Seyedi, F., Nejadghaderi, S. A., Sullman, M. J. M., Mohammadinasab, R., Kolahi, A. A., Arshi, S., & Safiri, S. (2023). Human monkeypox: history, presentations, transmission, epidemiology, diagnosis, treatment, and prevention. *Front Med*, 10, 1157670. https://doi.org/10.3389/fmed.2023.1157670

STUDIA UNIVERSITATIS BABEŞ-BOLYAI BIOLOGIA, 71, 1, 2026 (p. 101-124)
doi:10.24193/subbbiol.2026.1.06

Leaf morphological variation in common yew (*Taxus baccata* L.) at the species' southern range margin in North Africa

Chafiaa Hamidouche-Si Mohammed[1]✉, Tassadit Dib[1], Fazia Krouchi[1], Badri Wadi[2] and Arezki Derridj[1]

[1]Laboratory of Ecology, Biotechnilogy and Health, Mouloud Mammeri University, Tizi-Ouzou, Algeria; [2] University Hassan II of Casablanca, Faculty of Sciences Ben M'Sik, Laboratory of Ecology and Environment, Casablanca, Morocco, P.O.Box 7955;
✉***Corresponding author, E-mail: chafiaa.hamidouche@ummto.dz***

Article history: Received 8 January 2026; Revised 29 April 2026; Accepted 16 June 2026; Available online 25 June 2026

Abstract. The main aim of this study was to provide the first morphometric characterization of *Taxus baccata* L. leaves in North Africa. Phenotypic variability was investigated in nine natural populations from Algeria and Morocco using 11 quantitative and qualitative traits. analysis of variance (ANOVA) followed by Fisher's post hoc test revealed significant differences at multiple hierarchical levels, including trees within populations, among populations, regions and mountain ranges for all quantitative variables. In parallel chi-square analysis indicated significant variation in qualitative traits. Hierarchical clustering classified the studied populations into two major clusters for both types of morphological traits. The Chrea population displayed the largest leaves, whereas the Saharan Atlas populations showed the smallest. Principal component analysis (PCA) clearly separated populations and geographic groups in relation to both ecological and leaf morphological factors. Overall, the observed morphological differentiation suggested that phenotypic plasticity together with possible genetic differentiation, is driven more by local ecological conditions than by latitude alone. North African yews are characterized by relatively narow leaves and region specific variation in petiole and apex morphology, reflecting high

degree of ecological adaptability to various ecological and geographical conditions. These findings have important implications for conservation of southern marginal population of this relic threatened conifer.

Keywords: *Taxus baccata L.*, leaf morphology, morphometric variability, North Africa.

Introduction

The geographical range of *Taxus baccata L.* is extensive, spanning a broad latitudinal gradient from Scandinavia in the north margin to North Africa in the south, as well as a longitudinal gradient from England in the west to western Asia (Iran) in the east (Lapie and Maige, 1914; Maire, 1916; Thomas and Polwart, 2003). At the southernmost Mediterranean margin of its distribution, yew is limited to Algeria and Morocco, where it occurs in the understory of cedar and oak forests (Barbero and Quézel, 1994; M'hirt *et al*., 1999; Boudy, 1950).

Given its wide distribution range, *T. baccata* is exposed to diverse environmental conditions (Iszkuło *et al*., 2009; Cedro *et al*., 2011; Robakowski *et al*., 2018). In North Africa, the climate is characterized by hot, dry summers and humid to per-humid winters. In Algeria in particular, where *T. baccata* occupies a wide geographical range, it experiences pronounced climatic variability, from humid and per-humid bioclimates in the north to semi-arid to sub humid conditions in the south (Abdessemed, 1981; Derridj, 1990; Meddour, 2014). It can therefore be hypothesized that populations in these regions have developed specific adaptations to cope with such environmental constraints. Various plant traits, including leaf morphology and anatomy, may serve as indicators of these adaptations (Portsmuth *et al*., 2007; Wyka *et al*., 2007; Iszkuło *et al.*, 2012).

Plant responses to environmental variability involve multiple biological mechanisms, including growth (Angert, 2006; Iszkuło *et al*., 2012; Perrin *et al*., 2013), survival (Schirone *et al.,* 2010; Linares *et al.,* 2013; Núñez-Cruz et al., 2018), and reproduction and dispersal (Garcia *et al*., 2000; Iszkuło *et al*., 2009; Nkosi *et al*., 2024). These processes are often associated with variations in morphology, anatomy, physiology, and phenology, which ultimately influence the adaptive capacity of individuals to specific environments (Medina et al., 1990; Schirone *et al*., 2010; Zandalinas *et al.,* 2017). Leaves are particularly sensitive organs to the environmental conditions, and their morphological variability in size and shape is considerable, providing information on plant responses to habitat conditions (Boeger *et al.,* 2004; Boratyński, 2008; Pescador *et al.,* 2015).

Numerous studies on *T. baccata* leaf morphology have been conducted in southern Asia (Spjut, 2007; Shah *et al*., 2008), Europe (Mitchell, 1998; Schirone *et al.*, 2010; Vessella *et al*., 2013), and Siberia (Stefanovic *et al.*, 2017), highlighting substantial variability in leaf traits. Previous studies in the Azores (Schirone *et al*., 2010) and Madeira (Portugal) (Vessella *et al.,* 2013) identified the presence of minor morphotypes. These findings suggest that *T. baccata* populations in Macaronesian islands may represent relict lineages, with the occurrence of smaller leaves potentially resulting from long-term geographic isolation.

Despite increasing research on *Taxus baccata* L. across Europe and the Mediterranean Basin, populations in North Africa remain largely neglected. To date, only a few studies have addressed this region: Romo et al. (2017) documented the regression of yew in Morocco, Vessella *et al*. (2010) included three Algerian samples in a broader morphometric analysis of leaf traits, and Hamidouche *et al.* (2014) carried out a preliminary assessment of geographic distribution and general morphological variation in Algerian yew. However, no comprehensive investigation has yet targeted North Africa populations. Given the wide climatic variation across North Africa, it is expected that local populations exhibit specific leaf traits reflecting adaptation to these environments. This study addresses a critical knowledge gap on the adaptive and functional significance of leaf traits in marginal yew population. Morphometric analysis was employed to characterize *yew* populations across North Africa. A total of eight Algerian populations were investigated, including six from the Tell Atlas and two from the Saharan Atlas, along with one Moroccan populations from the Middle Atlas. This study aims to address the following questions: (i) What is the extent of variation in leaf morphology among North African yew populations? and (ii) Does leaf morphology in *T. baccata* vary along a latitudinal gradient in North Africa at its southern range margins?

Materials and methods

Study sites

Taxus baccata L., commonly known as yew, reaches the southern Mediterranean margin of its distribution in North Africa, where it occurs in the understory of *Quercus canariensis* forests and mixed *Cedrus atlantica–Quercus rotundifolia* stands. Its distribution in this region has been well documented in early floristic works (cited before) and further supported by phytosociological studies (M'hirt *et al.*, 1999; Laribi, 2000; Yahi, 2007).

Given the threatened status of *Taxus baccata* L. in North Africa, the sampling strategy was designed to be exhaustive as possible as at population level to capture the widest possible range of morphological variation within

each population. Adult trees were selected randomly among accessible individuals in order to ensure representative and unbiased sampling of the local population structure.

The present study included nine sites in North Africa, including eight in Algeria and one in Morocco (Fig. 1). In Algeria, sampling was conducted in mountainous areas of the Tellian and Saharan Atlas, between 1300 and 2500 m a.s.l. In the Tellian Atlas, three regions comprising six sites were investigated: Chréa (CHR), Djurdjura (Tala Guilef (TG), Tikjda (TK), Ath Ouabane (AOB), Ath Zikki (AZ)), and Akfadou (stand named Fontaine des Ifs) (AK). In the Saharan Atlas, populations were sampled in the Aurès Mountains, specifically at Djebel Chelia (CHL) and Djebel Feraoun (FER) The study also included one Moroccan site located in the Middle Atlas (Ifrane (IFR)). Figure 1 shows their geographic distribution, and Table 1 summarizes the ecological characteristics of the sites.

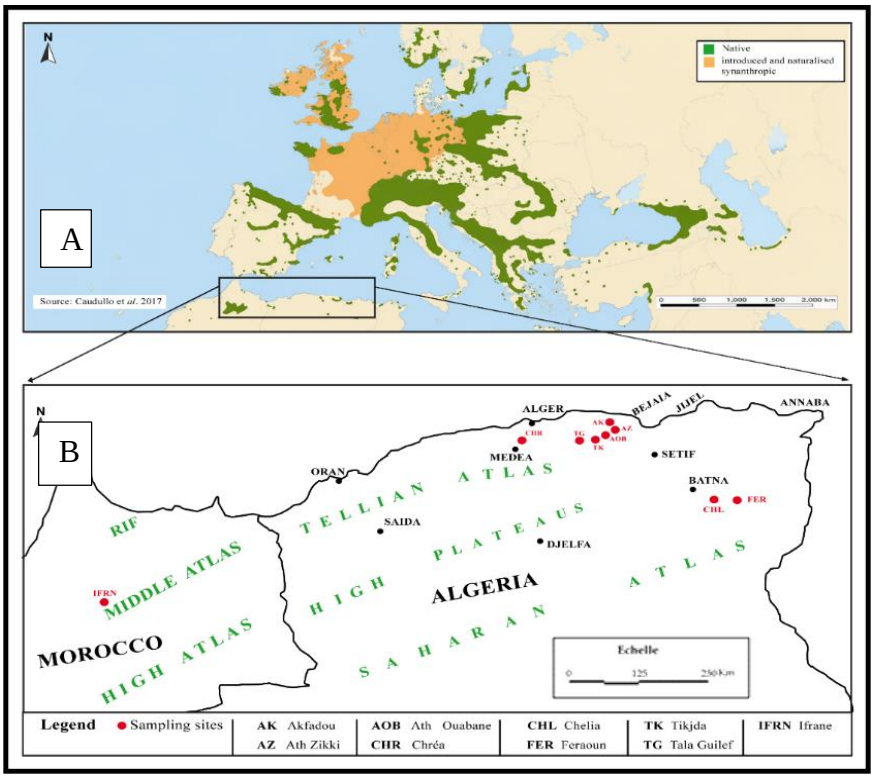


Figure 1. A-Word Geographic distribution of *Taxus baccata* L., B-Sampled population in North Africa (red dots).

Table 1. Geographic and ecological characteristics of the studied populations and sample size.

MOUNTAIN RANGE	**TALLIAN ATLAS**						**SAHARAN ATLAS**		**MIDLE ATLAS**
Region	**Djurdjura**				**Akfadou**	**Blidean Atlas**	**Aures**		**Morocco**
Stand name	**Tikjda (TK)**	**AthOuabane (AOB)**	**Tala Guilef (TG)**	**Ath Zikki (AZ)**	**Akfadou (AK)**	**Chrea (CHR)**	**Dj. chelia (CHL)**	**Dj. feraoun (FER)**	**Ifrane (IFR)**
Longitude	4.1263E	4.2925E	4.0007E	4.5088E	4.5739E	2.7211E	6.5647E	7.0957E	5.0972W
Latitude	36.4472N	36.4705N	36.4738N	36.5544N	36.6641N	36.3948N	35.3875N	35.3517N	33.5550N
Elevation (m)	1450	1000-1600	1300-2500	1500-1600	1130-1300	1300-1450	1983-2095	1830-1916	1400 -2080
Orientation	NW	N	N	N	N	NW	N and NW	NE	N
Slope	Low	Medum to High	High	High	Low	Low to medium	Medium to hiegh	Medium to hiegh	Medium to hiegh
Substrate	Calcareous	Calcareous	Calcareous	Calcareous	Numidian sandstone	Schist	White sandstone	Calcareous and stone	Calcareous
T (°c)	14,66	11,2	13,18	14,67	12,62	16,73	14,25	10,88	11
P (mm)	899,8	1169	939,4	900,15	987,01	676,92	363,58	483,6	647
Bioclimate	Humid	Humid	Humid	Humid	Subhumid to humid	Humid	Semiarid to subhumid	Semiarid to subhumid	Humid
Type of vegetation cover	*Cedrus atlantica, Quercus rotundifolia*	*Cedrus atlantica, Quercus rotundifolia*	*Cedrus atlantica, Quercus rotundifolia*	Remnant *Quercus rotundifolia*	*Quercus canariensis*	*Cedrus atlantica, Quercus rotundifolia*	*Cedrus atlantica*	*Cedrus atlantica*	*Cedrus atlantica*
Canopy %	60	60	70	10	80	90	65	50	
Sampled trees	8	14	13	3	10	11	10	10	10
Sampled leaves	240	560	390	90	300	330	300	300	300

Leaf morphology

Biological material was collected from adult *Taxus baccata* L. individuals, in total, 89 Taxus baccata trees were sampled, and leaves were collected from the mid-crown. for each tree, 30 to 40 undamaged leaves aged 2–3 years were collected and stored at 4 °C until measurement. Preliminary analyses were performed to test for potential variation among leaf orientations within individual trees (north, south, east, and west). As no significant differences were detected, leaf orientation was excluded from subsequent analyses in agreement with (Hamidouche et *al.,* 2014). The focus was placed on the accessible and sunny part of the crown. In total of 2810 leaves were collected, pressed and scanned with the abaxial surface facing upward using an Epson GT-15000 scanner at a resolution of 300 dpi. Size and shape were measured on image analysis using Digimizer 3.0 software. Five quantitative traits were measured (leaf area, leaf length, leaf width, petiole length, and apex length) and two derived ratios were calculated (leaf length/leaf width and leaf length/petiole length). The shape was further characterized using qualitative traits namely leaf curvature, leaf base, apex shape, and leaf apex (Fig. 2). These characters were assessed according to the previously published methodologies (Möller *et al*., 2007; Shah *et al.,* 2008; Vessella *et al.,* 2010).

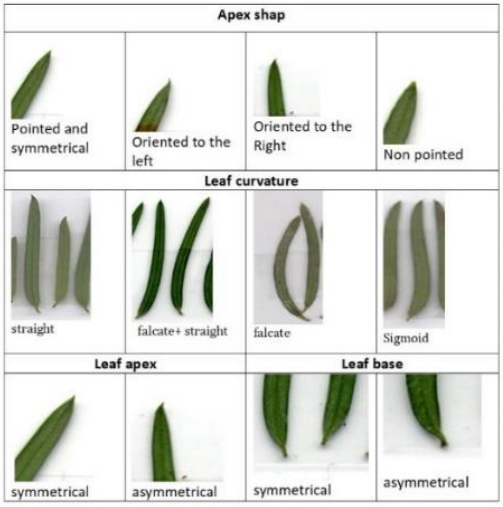


Figure 2. Assessment of qualitative traits of *Taxus baccata* L. leaves

Statistical analyses

Statistical analyses were carried out using **R** (version 4.0.3) and Excel software. The first step involved calculating descriptive statistics for each site. These statistics included the mean, the coefficient of variation and the range (from minimum to maximum).

To determine the differences between qualitative trait categories among populations, contingency tables were used. For each character a Pearson's Chi-square test (χ^2) of independence was performed

To explore the data structure, box plots were used to visualize the distribution of the measured variables.

To assess the variability between populations and among trees within the same population, across region as well as among Mountain ranges, analysis of variance (ANOVA) as normality of data and homogeneity of variances were confirmed, and Fisher's LSD post hoc test ($p< 0.05$) were performed separately for each level. Then a principal component analysis (PCA) was carried out to confirm patterns of similarity among yew populations and identify the main ecological and morphological factors likely to contribute to the convergence or divergence of populations, regions and mountain ranges.

To summarize the overall similarity among populations, Hierarchical clustering was used. For quantitative traits, based on standardized mean values, whereas qualitative variables clustering was applied to the frequency profiles of trait modalities observed withing each population, this approach provided a synthetic visualization of population affinities and completed the inferential analyses performed on individual traits

Results

Leaf shape variation

Analysis of qualitative traits related to Leaf shape variation revealed significant geographic diferences (χ^2 test, $p < 0.001$) for three traits (leaf curvature, leaf apex, and apex shape), wheras leaf base appeared relatively stable across populations (Tab. 2).

At population level, data reveal that sigmoid curvature was the rarest modality, whereas, the proportion of straight leaves ranged from 46.67% and 68.33%. The symmetrical apex was represented at rates above 58% among 05 populations. However, Tikjda and Akfadou populations were distinguished by a more balanced frequency between the two modalities. Furthermore, the apex shape indicated predominance of leaves with pointed and symmetrical apices

in eight of the nine studied populations whereas, Ath Ouabane population displayed equal proportions of pointed, symmetrical apices and non-pointed apices (Tab. 2).

Table 2. Frequencies of qualitative variables of *Taxus baccata* and Khi2 test results per variable.Leaf curvature: straight (St); falcate to straight (FSt),falcate (F), sigmoid (Sg); Apex shape : symmetrical (Sy), symmetrical (Asy); leaf base: symmetrical (Sy); Asymmetrical (Asy) and leaf shape: pointed and symmetrical (PS), pointed and oriented to the left(PL), pointed and oriented to the right(PR) and non- pointed(NP). population abbreviations follow Table 1.

Population	AK	AZ	CHR	TK	TG	AOB	CHL	FER	IFR
Leaf Curvature %	**(X^2_{obs} = 244,7; X^2 th = 51,18; α =1‰)**								
St	52,33	46,67	57,88	47,08	57,18	59,11	50,00	55,67	68,33
FSt	16,00	13,33	15,15	13,33	17,18	22,86	34,00	27,33	25,00
F	18,67	40,00	24,24	29,58	18,21	11,07	14,67	14,00	5,33
Sg	13,00	0,00	2,73	10,00	7,44	6,96	1,33	3,00	1,33
Leaf base %	**(X^2_{obs} = 31,15; X^2 th = 51,18; α =1‰)**								
SY	30,67	46,67	32,12	34,58	35,64	37,50	41,67	39,33	25,33
SY	69,33	53,33	67,88	65,42	64,36	62,50	58,33	60,67	74,67
Leaf Apex %	**(X^2_{obs} = 401,80; X^2 th = 51,18; α =1‰)**								
sy	40,00	63,33	57,88	49,17	75,13	78,57	90,00	92,33	85,67
Asy	60,00	36,67	42,12	50,83	24,87	21,43	10,00	7,67	14,33
Apex shape %	**(X^2_{obs} = 449,36; X^2 th = 51,18; α =1‰)**								
PS	36,00	63,33	58,48	48,33	57,69	39,11	80,67	77,67	80,00
PL	19,33	15,56	14,85	17,50	9,74	12,50	6,00	5,33	12,00
PR	19,66	10,00	4,85	16,25	10,00	8,57	7,33	7,33	7,00
NP	26,00	11,11	21,82	17,92	22,56	39,82	6,00	9,67	1,00

Leaf size variation

The descriptive statistics of *quantitative leaf morphology* revealed variations both among the populations and within populations for the studied traits, highlighting differences among trees, as reflected by ranges between minimum and maximum values (Table 3).

Boxplot showed a clear interpopulation differenceswhile the presence of the Outliers, reflected intrapopulation variability (Fig. 3). within each populations some trees are displayed high values for all the studied traits, whereas others showed consistently lower values, this pattern was observed within all the populations, supporting the significant intra-population variability among in leaf morphology.

The overall leaf area mean registered in our study was 42,54 mm^2. The Chrea population exhibited the largest leaf areas with 90% of needles exceding 50 mm^2, whereas the Saharan Atlas populations displayed the smallest areas, with most leaves measuring less than 40 mm^2 (data non shown). Leaf length varied markedly among populations ranging from a mean of 22.59 mm in FER to 29.0 mm in CHR. In contrast, leaf width was less variable, with an overall mean of 2.15 mm and cv=12,47. Population means ranging from 1.96 mm inTK to 2.39 mm in CHR populations.

The calculated L/W ratio averaged 10.47 and ranged from 9.69 in CHL to 12.19 in CHR sites. the greatest petiole lengths were observed in TG and IFR sites (1.28 mm) and the shortest petioles were recorded in the Saharan Atlas populations (0.97-1mm). The calculated L/Pl ratio averaged 20.37, with the lowest mean in AOB and AZ and the highest in Chrea sites (25.32). Finally, apex length ranged from 0.32 mm in FER to 0.53 mm in AK .

The performed ANOVA analysis revealed highly significant differences across populations for all studied quantitative traits with distinguished homogenous groups according to Fisher s test post hoc (Fig. 4). it appeared that chrea formed the first group for all trait except petiole lenght foe whish Ath Zikki exhibited the highest values and apexe lenght for whish Akadou was ranked first.

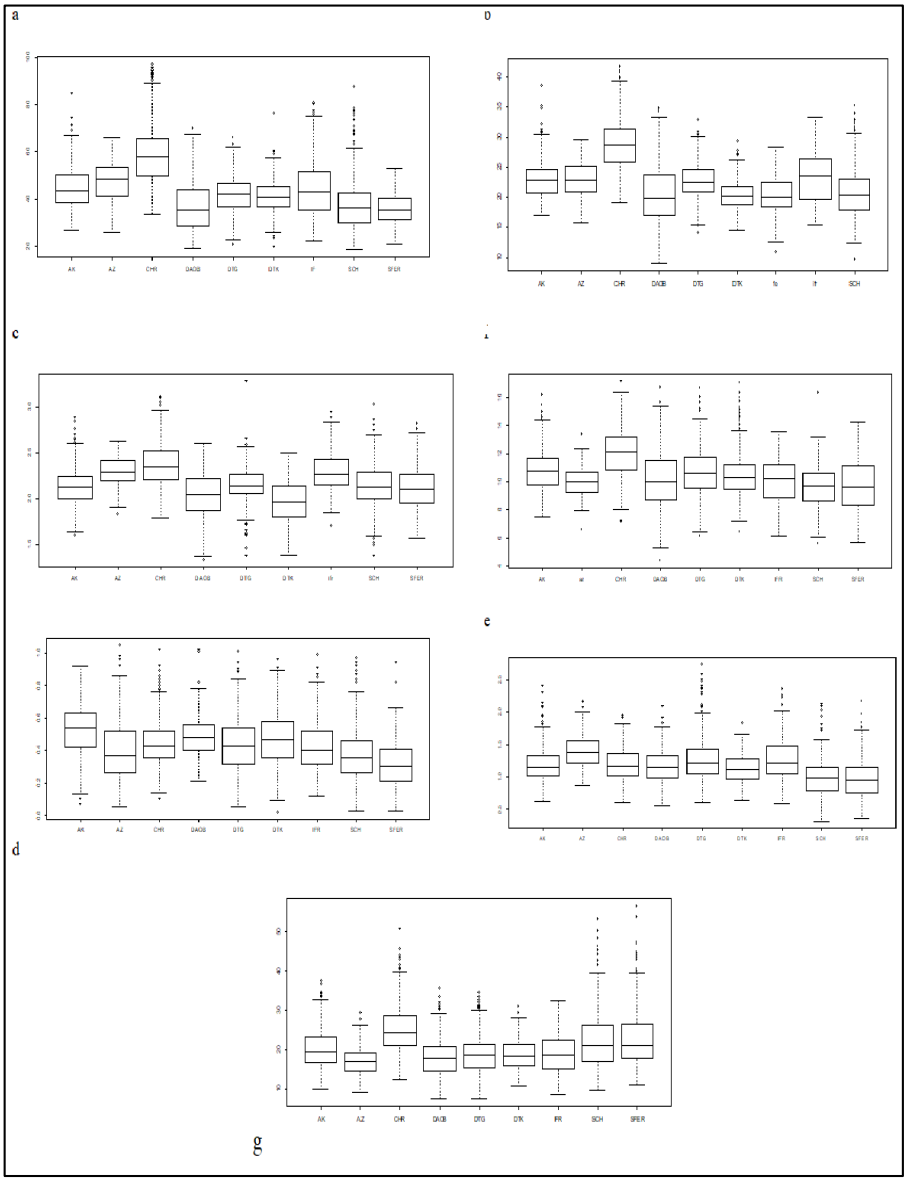


Figure3. Mean and range values distribution of *T. baccata* per site. Morphometric traits -a-leaf area; b-leaf length; c-leaf width; d- petiole length, and e-apex length; calculated ratio: f-leaf length/leaf width (L/W) and g-leaf length/petiole Length, population abbreviations follow Table 1.

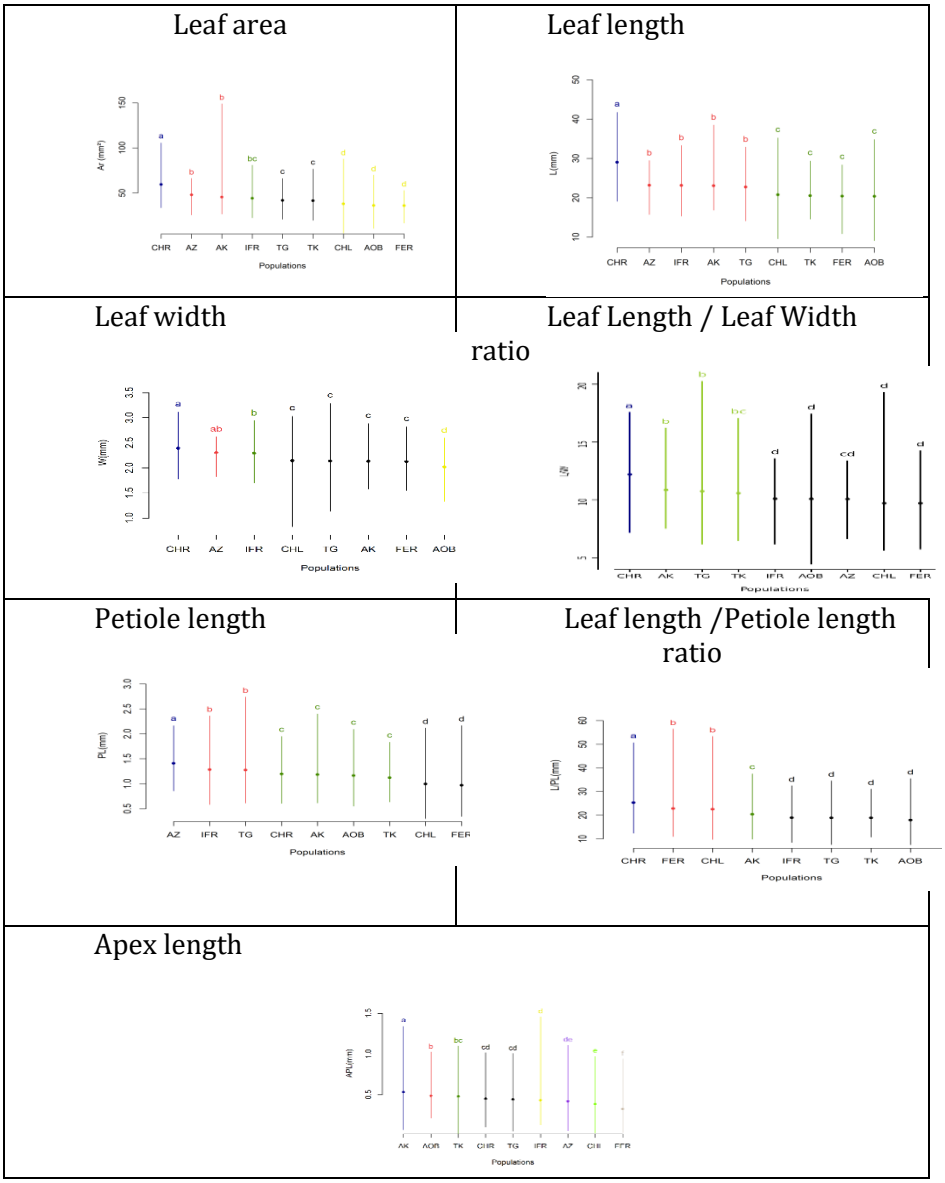


Figure 4. Leaf morphology variation among *Taxus baccata* populations and Fisher's homogeneous groups. Population abbreviations follow Table 1

Table 3. Mean values (mean ± standart errors), homogenous groups according to Fisher's LSD test ($p < 0.05$); -coefficient of variation (cv %) and Minium-Maximun values (Min-Max) of studied *T. baccata* populations population abbreviations follow Table 1.

Mountain range	*Region*	*Population*	*Tree (N)*	*Leaf (N)*	*Statistical parameters*	Leaf area *(Ar) (mm2)*	Leaf Length *(L) (mm)*	leaf width *(W) (mm)*	Leaf Length/ width *(L /W)*	petiole length *(Pl) (mm)*	Leaf length /petiole length *(L/Pl)*	Apex length *(Apl) (mm)*
Tellian Atlas	Djurdjura	**TG**	13	390	mean ± se	41,87±0,60 c	22,74±0,19b	2,14 ±0,01 c	10,75±0,09 b	1,27±0,01 b	18,86±0,28 d	0,44±0,007 cd
					cv	18,71	13,6	10,58	16,66	27,36	24,75	34,62
					Min-Max	20,88-66,13	14,09-32,89	1,14-3,28	6,13-20,26	0,61 -2,73	7,55-34,57	0,05-1,00
		TK	8	240	mean ± se	41,47±0,67c	20,51 ±0,24 c	1,96±0,01 e	10,59±0,12bc	1,12±0,01 c	18,86±0,35 d	0,48 ±0,009 bc
					cv	18,4	12,55	10,95	17,28	19,32	19,69	36,53
					Min-Max	19,66-76,53	14,55-29,33	1,37-2,50	6,46-17,07	0,64-1,83	10,71-31,08	0,02 -1,09
		AOB	14	560	mean ± se	36,25±0,44 **d**	20,36 ±0,16 c	2,02±0,01 d	10,09±0,09d	1,16± c	17,98±0,23 d	0,48 ±0,006 b
					cv	30,61	22,47	11,57	20,06	20,07	25,08	24,4
					Min-Max	10,67 - 69,94	9,10-34,80	1,33-2,59	4,40-17,44	0,55-2,09	7,49 -35,51	0,21-1,02
		AZ	3	90	mean ± se	47,83±1,11 b	23,13±0,4 b	2,3±0,02 ab	10,06±0,12cd	1,41±0,03 a	17,06±0,58 d	0,42±0,016 de
					Cv	17,34	11,97	7,2	11,54	20,34	22,9	52,88
					Min-Max	25,75-66,21	15,74 -29,45	1,83-2,62	6,59 -13,4	0,85 -2,16	8,99 -29,32	0,05-1,11
	Akfadou	AK	10	300	mean ± se	45,47±0,60 b	23,07±o,21 b	2,13± 0,01 c	10,85 ±0,09b	1,18 ±0,01 c	20,39 ±0,32 c	0,53 ±0,008 a
					Cv	25,32	15,12	10,56	13,72	24,38	24,44	30,04
					Min-Max	26,63-149,04	16,85-38,49	1,58-2,87	7,50 -16,21	0,62-2,39	9,83 -37,48	0,07-1,34
	Chrea	**CHR**	11	330	mean ± se	59,55±0,57 a	29±0,20 a	2,39±0,01 a	12,19±0,01 a	1,19 ±0,01 c	25,32±0,30 a	0,45 ±0,008 cd
					Cv	30,61	22,47	11,57	20,06	20,07	25,08	24,4
					Min-Max	33,71-105,52	19,08-41,70	1,77-3,11	7,15-17,56	0,61-1,94	12,42-50,66	0,11-1,02
Saharan Atlas	Aures	**CHL**	10	300	mean ± se	37,8±0,60 d	20,71±0,21 c	2,14 ±0,01 c	9,69±0,09d	1±0,01 d	22,54±0,32 b	0,38 0,008 e
					Cv	31,5	19,33	12,56	16,44	30,52	33,59	42,63
					Min-Max	5,57-87,71	9,60 -35,22	0,83 -3,03	5,59 -19,29	0,31-2,1 2	9,71-53,33	0,3 - 0,96
		FER	10	300	mean ± se	35,96±0,60 b	20,39±0,22 c	2,13 ±0,01c	9,7±0,11 g	0,97±0,01 d	22,8±0,32b	0,32±0,008 f
					Cv	18,37	15,02	10,79	18,76	30,37	32,87	45,14
					Min-Max	16,62-52,84	10,83 -28,36	1,55 -2,81	5,7 -14,25	0,34 -2,17	10,99 -56,48	0,03 -0,94
Middle Atlas	Ifrane	**IFR**	10	300	mean ± se	44,13±0,60bc	23,15±0,21b	2,29±0,01 b	10,09±0,09d	1,28±0,01 b	18,92±0,32 d	0,43±0,009 d
					Cv	25,85	17,15	9,65	14,62	24,41	25,79	41,09
					Min Max	22,53-80,98	15,33-33,30	1,69 -2,94	6,1-3,57	0,58-2,36	8,44-32,50	0,1-1,46

The same tendency was observed at regional scale, Chréa region exhibited highest values for leaf area, length, width L/W ratio, and L/Pl ratio, Ifrane showed the heighest petiole lenght followed by Djurdjura. And Akfadou desplayed the first rank for the apex lenght. On another hand Aures populations (CHL and FER) were characterized by the smallest leaf size (Fig. 5).

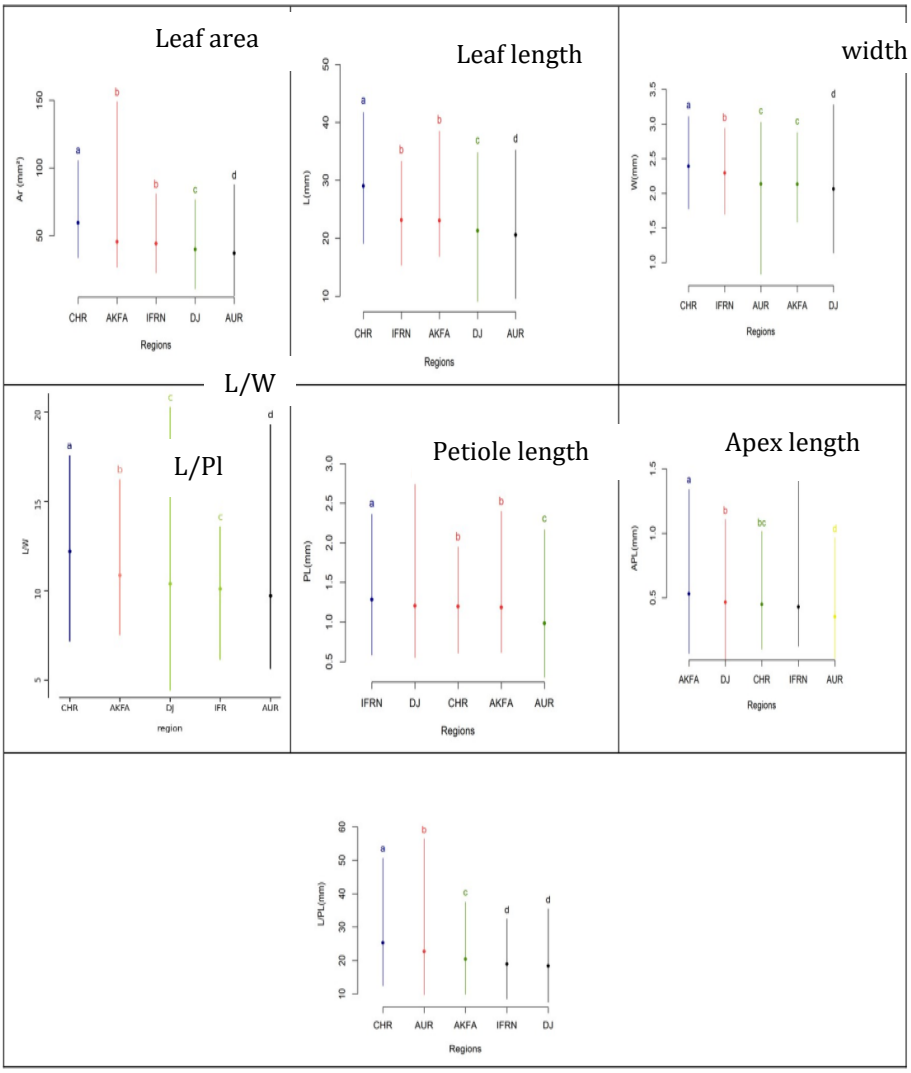


Figure 5. Leaf morphology variation among Taxus *baccata* region and Fisher's homogeneous groups. Population abbreviations follow Table 1

The comparaison of mean values among the three mountaine ranges revealed significant range variation in quantitaive leaf morphological traits. Populations from Tellian Atlas and Middle Atlas formed a homogeneous group for the leaf area and lenght, whereas they differed significatively in leaf width, petiole lenght and the two calculated ratios. By contrast populations from saharan Atlas showed the lowest values for the most traits except the apex lenght for whish the Saharan Atlas was assigned to the higher ranking group (Fig. 6).

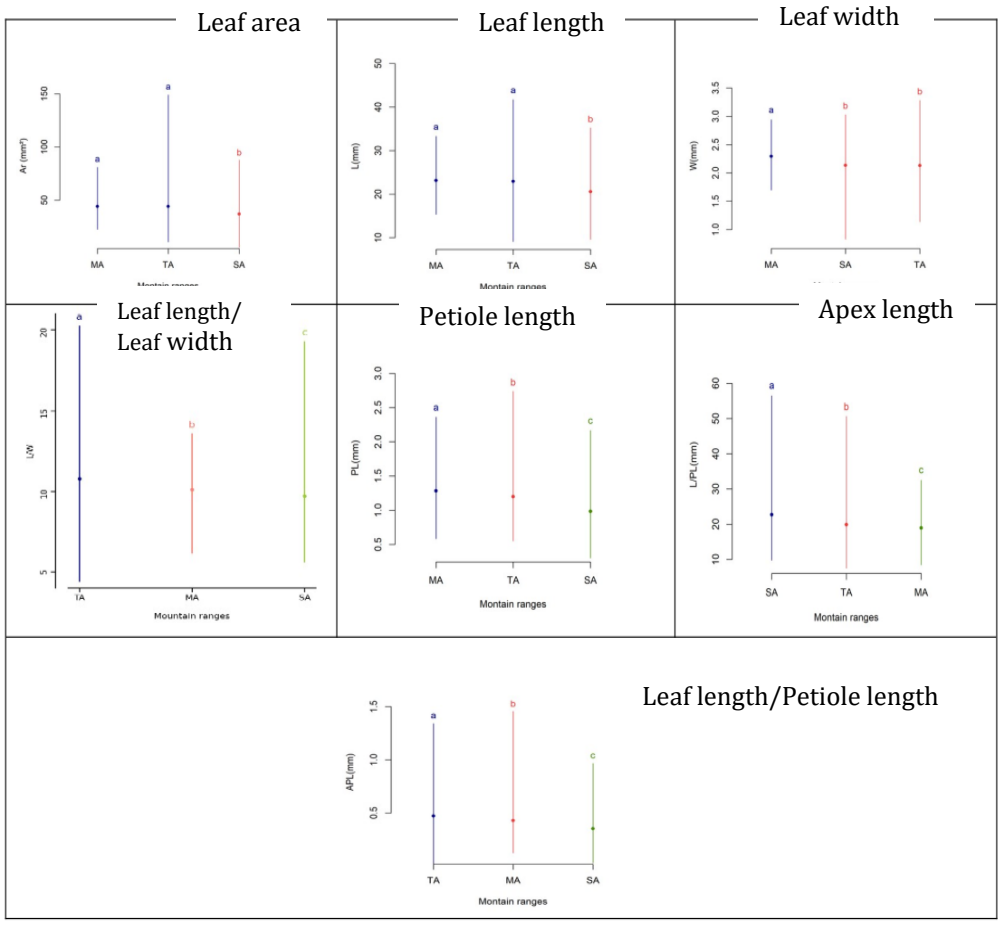


Figure 6. Leaf morphology variation among mountain ranges and Fisher's homogeneous groups.

Principal component analysis (PCA) explained 68.8% of the total variance across the first two axes (dimension 1: 42.7%; dimension 2: 26.1%). The first dimension primarily related to the bioclimate, altitude, and leaf morphological parameters, such as leaf area, leaf length, petiole length, and apex. The second dimension was associated with leaf width, annual precipitation, and leaf length (Fig. 7). The analysis clearly separated the Saharan Atlas populations from those of the Tellian Atlas, while the population of the Middle Atlas occupied an intermediate position. PCA correlation circle showed that latitude was positively associated with annual precipitation and negatively with the altitude.

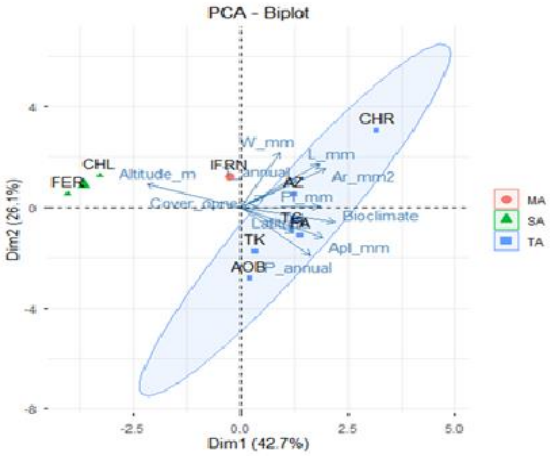


Figure 7. Principal component analysis performed for nine *Taxus baccata* L. populations based on morphometric, geographic and environmental data.

Cluster analysis revealed that North African yew populations were structured into two main groups according to leaf morphology, and corroborate with the results of ANOVA followed by Fisher's test and PCA results. for the quantitative trait and are consistent with the chi square test for the qualitative traits.

Two major groups were defined For quantitative variables. Chrea was the most divergent population being separated from the remaining populations. Within the second group, The Saharan Atlas population were closely associated, while Middle and Tell Atlas populations showed substantial similarity (Fig. 8A).

Also, two distinct groups were illustrated on the qualitative characters. Ifrane from Midlle Atlas clustered with geographically close Saharan Atlas populations, while the Tell Atlas populations showed affinities among themselves as well as AOB with TG, AK with TIK and CHR with AZ (Fig. 8B).

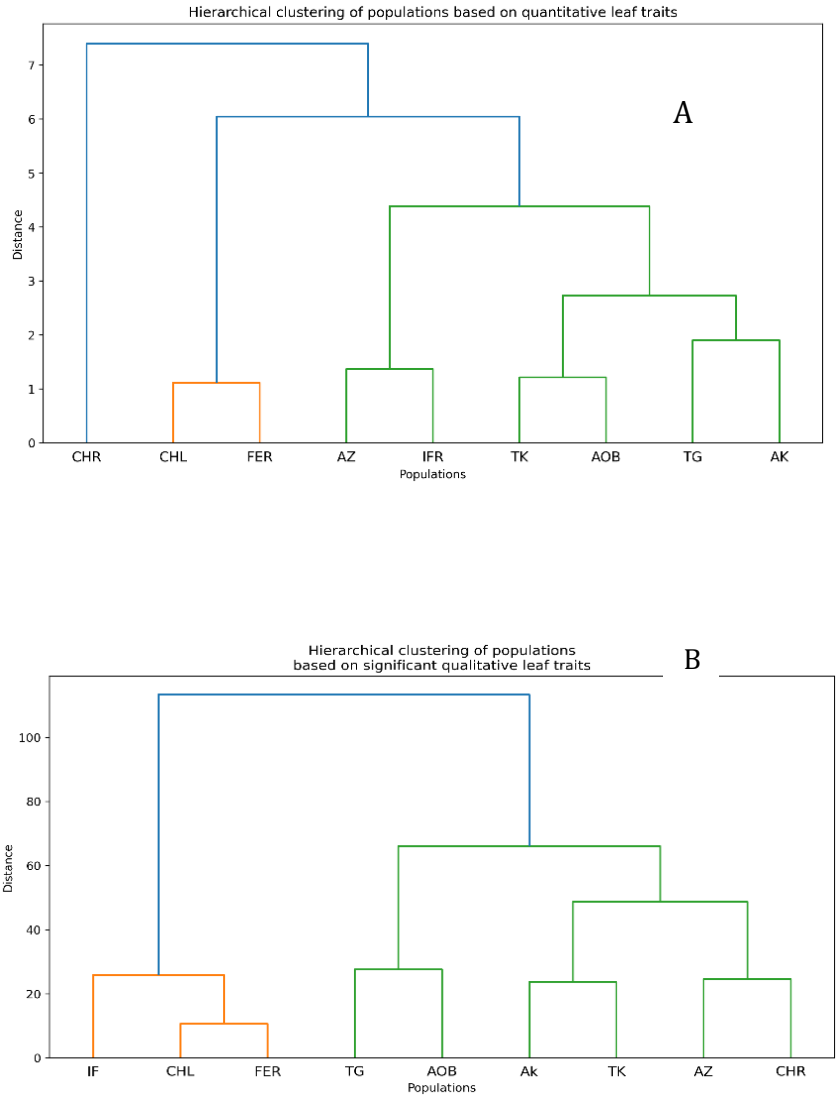


Figure 8. Hierarchical clustering based on leaf morphology characters. A-quantitative traits, B-qualitative traits.

Discussion

This study provides the first broad assessment of leaf morphological variation in *Taxus baccata* L. at its southern margin in North Africa and revealed substantial differentiation at multiple hierarchical levels. variation was recorded among populations, regions and mountain ranges and as well as among individual trees within populations, indicating that leaf morphology in North African yew was structured by both local and broader environmental factors. The relatively low standard errors associated with populations mean, together with wide coefficients of variation, indicated strong variability. Together, these results suggested that the joint influence of environmental heterogeneity and genetic differentiation.contributed to the observed morphological structure

First, an important result is the marked variation observed among individual trees within the same population. Such intra-population variability likely reflected the joint effects of intrinsic individual-tree characteristics, including age (Omarova, 2018), sex (Iszkuło et al, 2009) and vigour, together with local microenvironmental conditions related to topography, soil properties, canopy cover and light availability. Comparable tree level structuring of leaf morphology has been reported in *T. baccata* populations from the Balkans and Dagestan populations (Tumpa *et al*., 2022), in Siberian populations (Stefanovic *et al*., 2017), and in Himalayan *T. fauna* (Li *et al*., 2022), supporting the view that both phenotypic plasticity and genetic differentiation may operate at the tree level. Similar interpretations have also been advanced for other woody taxa, including *Quercus petraea* and Norway spruce, where variation among individuals has been linked to both intrinsic and local environmental conditions (Bruschi *et al.,* 2003; Metslaid et al., 2007 and Kenzo *et al.*, 2006).

At the population and regional scales, leaf morphometric traits clearly discriminated the main biogeographical groups. The Chréa population exhibited the largest leaves whereas populations from the Aurès consistently displayed the smallest one. Djurdjura and Moroccan populations were distinguished by longer petioles, while the Akfadou population showed longer apices. These differences, supported by both ANOVA and PCA analyses, indicated that leaf traits provide effective descriptors of population structure in *Taxus baccata* L. Hirarchical Clustering illustrated populations occurring at comparable environmental or morphological characteristics. More importantly, the multivariate analysis suggested that population differentiation followed an ecological gradient rather than a strictly geographical one. Higher latitude populations tended to occur under wetter climatic conditions, while lower latitude populations were associated with higher altitudes. Saharan Atlas populations were clearly separated from those of the Middle and Tell Atlas, and trait values generally decreased from the more humid northen sites toward the drier Saharan Atlas.

This, suggested that the observed variation among North African yew populations may be driven by local environmental adaptation, particularly water availability, rather than latitude alone.

The ecological contrast among sites helps explain these trends. Chréa appeared to represent the most favorable habitat, owing to its dense vegetation cover as. humid to perhumid bioclimate (Meddour, 2014) and proximity to the Mediterranean Sea, all of which contribute to cool and moist understory conditions suitable for yew persistence and growth. By contrast, the smaller leaves observed in the Aures populations were likely associated with prolonged summer drought (Beghami, 2010), more open canopies, and stronger disturbance due to anthropogenic disturbance, as well as ongoing habitat degradation linked to climatic and global changes (Abdessamed, 1990 Ramo *et al.,* 2017; Sarmoum, 2019). Such conditions increase sun exposure and reduce the maintenance of the humid microclimate required by *Taxus baccata* L., thereby imposing stronger water stress. In this context, reduced leaf size may represent a functional adjustment that limits transpirational water loss and improves drought resistance(Read et al., 2014). Ath Zikki mountain, constututes a particularly informative case:its yew individuals persist in deep, narrow roky fissures, where snow accumulates and humidity is retained over long priods (Maire, 1916 and Hamidouche *et al*. (2014).suggesting that this site functions as a local microrefugium buffering regional climatic stress.

These ecological interpretations are consistent with previous studies that describe leaf area as plastic trait closely linked to light interception and photosynthetic performance. Larger leaves are generally associated with shaded conditions, whereas smaller leaves tend to occur under high irradiance (Mitchell and Arnott, 1995; Perrin *et al.*, 2013 and Kostrakiewicz 2009). In the present study, the largest leaf areas were recorded in Chréa, Ath Zikki, Akfadou and Ifrane, which supported the hypothesis that local ecological conditions strongly influence leaf development. Under Shaded conditions in cedar and oak forests and specific habitat in Ath Zikki where the main ecological requirements for yew persistence are found, larger leaves likely enhance light capture and carbon gain. Conversely, in drier and more open habitats (Chelia and Feraoun), smaller leaves may reduce water loss .and improve stress tolerance.

More broadly, North African populations appear morphologicaly distinctive. Mean leaf area was higher than the values recorded for several Mediterranean and Atlantic European regions including Azores, Carpineto and Rossello, Madeira, Spain and Greece which reported values are 21.3, 33.99, 35.49 and 34.68 mm^2, respectively (Schirone *et al.,* 2010; Vessella et al., 2013), but lower than the values reported in Poland (57.24 mm^2) and in Serbia (47.18 mm^2) by Stefanović *et al.(*2017) and Zarek (2007). Likewise, mean leaf length was close to that reported for Siberian populations and exceeded values previously

recorded from some Atlantic and Mediterranean sites. By contrast leaf width in North Africa showed relatively lower values, resulting in a comparetively high leaf length-to-width ratio and indicating that North African yew populations tend to have narrower leaves. The combination of relatively elongated and narrow leaves may represent a characteristic feature of yew at its southern range margin. Petiole length and apex length also emerged as informative traits for regional discrimination suggesting that these components of leaf architecture are especially responsive to environmental gradients. These findings are consistent with those of Roche *et al.,* (2004) and Pasini *et al*. (2006) who showed that some leaf traits are more reliable than others for ranking plant species along an environmental gradient.

The present results support the view that bioclimate and elevation may exert a stronger influence than latitude alone. Although latitude is often invoked to explain large scale trait variation. The North African case shows that local ecological compensation may create suitable conditions for persistence even at the warm and dry southern edge of the *Taxus baccata* range. This interpretation is in line with previous studies showing that leaf traits are often shaped more strongly by local drivers such as moisture, temperature and canopy structure than by broad geographic position (Alcántara *et al*., 2000). Leaf size often decreases with increasing altitude (Thomas and Polwart, 2003; Iszkuło, 2006; Stefanovic *et al.,* 2017)

Qualitative traits provided additional support for strong morphological differentiation. Leaf base shape remained relatively constant among populations, whereas leaf curvature, apex symmetry, and apex shape varied significantly among sites. The predominance of straight leaves with asymmetrical bases broadly agrees with previous descriptions of *T. baccata* needles (Elwes and Henry, 1906, cited in Dempsey and Hook, 2000; and Shah *et al.,* 2008 and Stefanovic *et al*., 2017), although other authors have emphasized falcate or recurved needles in yews (Spjut, 2007). In our material, leaf apices were predominantly pointed and symmetrical. in contrast to Shah et *al.* (2008), who noted an indistinct mucro in *T. baccata*, and to Cope (1998), who reported a short petiole and an acuminate to cuspidate apex, often angled or flattened at the tip. Taken together, this qualitative variation further reinforces the view of high phenotypic plasticity and marked geographic differentiation in *T. baccata* foliage.

Overall, the univariate and multivariate analyses converge on the conclusion that North African populations of *T. baccata* L. exhibit pronounced morphological differentiation, at both intra population and inter population levels. This variability most likely reflects the interaction between local ecological conditions and population differentiation, and it should be explicitly considered in conservation strategies for this relict and threatened conifer.

Conclusion

In this study, we quantified variation in leaf morphology of *Taxus baccata* L. across North African populations, by analysing the size and shape of 2810 leaves among 89 trees, collected from eight populations in Algeria and one in Morocco. Leaf morphometric traits proved to be powerful descriptors, revealing pronounced morphological differentiation at tree, population, region and mountain range scales. Leaf traits effectively discriminated biogeographical groups and showed clear structuring along ecological gradients. The observed variability reflected the combined influence of local ecological conditions (microclimate, canopy structure and topography) and regional to mountain climatic gradients (humidity, drought intensity, elevation and latitude) and likely underlying epigenetic differentiations.

North African *Taxus baccata* L. exhibited distinctive leaf architecture, notably narrow leaves and region -specific petiole and apex morphology. At its southern range margin, yew represents edge populations exposed to warm and frequently dry conditions where favourable local conditions compensate for otherwise limiting macroclimate. These populations therfore represent a key reservoir of morphological and potential genetic diversity. Conservation strategies should therfore adopt a population and region oriented approach, preserving population across the full altitudinal and latitudinal distributions and maintaining the adaptive potential of this relict and threatened conifer under ongoing climatic and anthropogenic pressures.

Acknowledgements: Our thanks go to All the Algerian Forestry authorities, wilaya of Tizi Ouzou, Batna, Khenchela and National Parc of Djurdjura and Chrea) for their assistance in facilitating and contributing to the fieldwork.

We also thank all individuals who participated in data collection and to provide valuable help, during the field campaigns.

References

Abdessemed K. (1981). Le cèdre de l'Atlas dans les massifs de l'Aurès et du Belezma Etude phytosociologique et problèmes de conservation et d'aménagement [in French]. *Thèse Doc-Ing., Univ. Aix-Marseille III*, 199p.

Angert, A. L. (2006). Growth and leaf physiology of monkey flowers with different altitude ranges. *Oecologia*, *148(2),*183-194. https://doi.org/10.1007/s00442-006-061-z

Barbero, M. & Quézel, P. (1994). Place, rôle et valeur historique des éléments laurifoliés dans les végétations préforestières et forestières ouest-méditerranéennes [in French]. *Annal de Botanique Volume, 52*. 81-133.

Beghami, Y. (2010). Contribution à l'étude de l'influence des facteurs édaphiques, orographiques et biologiques sur la régénération naturelle du cèdre de l'Atlas (*Cedrus atlantica*. M) dans les monts des Ouled Yagoub [in French]. *These doc. Univ. de biskra*, Algerie 288 p

Boeger, M., Alves, L. & Negrelle, R. (2004). Leaf morphology of 89 tree species from a lowland tropical rain forest (Atlantic Forest) in South Brazil. *Brazilian Archives of Biology and Technology* 47. https://doi.org/10.1590/S1516-89132004000600013

Boudy, P. (1952). Guide forestier en Afrique du Nord. *La maison rustique*, Paris. 505 p.

Bresson, C. Vitasse, Y. Kremer, A. & Delzon, S. (2013). To what extent is altitudinal variation of functional traits driven by genetic adaptation in European oak and beech? *Tree Physiology. 31*, 1164-1174. https://doi.org/10.1093/treephys/tpr084

Bruschi, P., Grossoni, P. & Bussotti, F. (2003). Within- and among-tree variation in leaf morphology of *Quercus petraea* (Matt.) Liebl. natural populations. *Trees 17*, 164–172. https://doi.org/10.1007/s00468-002-0218-y

Caudullo, G., Welk, E. & San-Miguel-Ayanz, J. (2017). Chorological maps for the main European woody species. *Data in Brief, 12*, 662-666. https://doi.org/10.1016/j.dib.2017.05.007 .

Cedro, A. & Iszkuło, G. (2011) Do females differ from males of european yew (taxus baccata l.) in dendrochronological analysis? *Tree-Ring Research*, *67*(1), 3-11 https://doi.org/10.3959/2009-9.1

Cope, S., Corney, D., Clark, Y., Remagnino,P., Wilkin, P. & Debeaux O. (2012); Plant species identification using digital morphometrics: *A Review Expert Systems with Applications 39*, 7562–7573. https://doi.org/10.1016/j.eswa.2012.01.073

Dempsey, D. & Hook, I. (2000). Yew (*Taxus*) species - Chemical and morphological variations. *Pharmaceutical Biology 38*(4), 274-280. https://doi.org/10.1076/1388-0209(200009)3841-aft274

Derridj, A. (1990). Etude des populations de *Cedrus atlantica* Manetti en Algérie[in French]. *Thèse Doc., Univ. Paul Sabatier de Toulouse*, 288 p.

Garcia, D., Zamora, R., Hódar, J. A., Gómez, J. M. & Castro, J. (2000). Yew (*Taxus baccata* L.) regeneration is facilitated by fleshy-fruited shrubs in Mediterranean environments. *Biological Conservation, 95*, 31-38. https://doi.org/10.1016/S0006-3207(00)00016-1

Hamidouche-Si, M. C., Krouchi, F., Bouamed, A. & Vessella, F. (2014). Geographic distribution and morphological variation of *Taxus baccata* in Algeria. *Der Eibenfreund, 20*, 39–53.

Iszkuło, G., Jasińska, A., Giertych, M. & Boratyński, A. (2009). Do secondary sexual dimorphism and female intolerance to drought influence the sex ratio and extinction risk of *Taxus baccata*? *Plant Ecology 200*(2), 229-240. https://doi.org/10.1007/s11258-008-9447-5

Iszkuło, G. & BoratyńskI, A. (2006). Analysis of the relationship between photosynthetic photon flux density and natural *Taxus baccata* seedlings occurrence. *Acta Oecologica*, *29*, 78 84. https://doi.org/10.1016/j.actao.2005.08.001

Iszkuło, G., Didukh, Y., Giertych, M. J., Jasińska, A. K., Sobierajska, K. & Szmyt, J.(2012). Weak competitive ability may explain decline of *Taxus baccata*. *Annals of Forest Science, 69*, 705-712. https://doi.org/10.1007/s13595-012-0193-4

Kenzo, T. (2012). Variations in leaf photosynthetic and morphological traits with tree height in various tree species in a Cambodian tropical dry evergreen forest. *Japan Agricultural Research Quarterly, 46*, 167-180. https://doi.org/10.6090/jarq.46.167

Kostrakiewicz, K. (2009). The influence of shadow created by adjacent plantson phenotypic plasticity of endangered species *Trollius europaeus* l (Ranunculacae). *Polish Journal of Ecology: 57*(4), 625-634.

Lapie G. & Maige A. (1914). Flore forestière illustrée, comprenant toutes les espèces ligneuses de l'Algérie et les espèces ligneuses les plus répandues en Tunisie, au Maroc et dans le Midi de la France[in French]. *Édition Orlhac E., Paris*. 357p.

Laribi, M. (2000). Contribution à l'étude phytosociologique des formations caducifoliées à *Quercus canariensis* Willd et *Quercus afares* Pomel du massif forestier d'Ath-Ghobri et d'Akfadou (Kabylie) [in French]. *Mémoire de Magister, Universite Mouloud Mammeri, Tizi ouzou,* 140p.

Li, T., Shen-Tu, X., Xu, L., Zhang, W., Duan, J., Song, Y. & Dong, M. (2022). Intraspecific and sex-dependent variation of leaf traits along altitude gradient in the endangered dioecious tree *Taxus fuana* Nan Li & R.R. Mill. *Frontiers in Plant Science, 13*, 996750. https://doi.org/10.3389/fpls.2022.996750

Linares, J. C. (2013). Shifting limiting factors for population dynamics and conservation status of the endangered English yew (*Taxus baccata* L., Taxaceae). *Forest Ecology and Management*, *291*, 119-127. https://doi.org/10.1016/j.foreco.2012.11.009

Maire R., (1916).Deuxième contribution à l'étude de la flore du Djurdjura. *Bulletin de la Société d'Histoire Naturelle de l'Afrique du Nord*, *Tome. 07*(2), 49-61.

Mayol,M., Riba, M., Gonzalez-Martınez, S., Bagnoli,F., de Beaulieu, J., Berganzo, E., Burgarella, Dubreuil, M., Krajmerov D., Paule, L., Romsakova, I., Vettori, C., Vincenot, L. & Vendramin, G. (2015). Adapting through glacial cycles: Insights from a long-lived tree (*Taxus baccata*). *New Phytologis, 208*, 973–986. https://doi.org/10.1111/nph.13496

Meddour R. (1994). La cédraie de l'Atlas Blidéen (Algérie): Valeur bioclimatique, syntaxonomique et dynamique[in French]. *Annals de Recherches Forestieres, Maroc, T. 27*(1), 105-27.

Medina, E., Garci, V. & Cuevas, E. (1990). Sclerophylly and oligotrophic environments : relationships between leaf structure, mineral nutrient content, and drought resistance in tropical rain forests of the upper rio negro region. *Biotropica, 22*(1), 51-64. https://doi.org/10.2307/2388719

Metslaid, M., Jogiste, K., Nikinmaa, E., Moser, W. & Porcar-Castell, A. (2007). Tree variables related to growth response and acclimation of advance regeneration of *Norway spruce* and other coniferous species after release. *Forest Ecology and Management: 250,* 56–63. https://doi.org/10.1016/j.foreco.2007.03.009

M'hirit O. & Blerot P. (1999). Le grand livre de la forêt marocaine[in French]. *Edition Mardaga,* 280 p.

Mitchell, A. K. & Arnott, J.T. (1995). Effects of shade on the morphology and physiology of *amabilis fir* and western hemlock seedlings. *New Forests 10*, 79-98. https://doi.org/10.1007/BF00034177

Mitchell, A. K. (1998). Acclimation of Pacific yew (*Taxus brevifolia*) foliage to sun and shade. *Tree Physiology 18*, 749-757. https://doi.org/10.1093/treephys/18.11.749

Moller M., Gao, L., Mill, R., Li, D., Hollingsworth, M. & Gibby, M. (2007). Morphometric analysis of the *Taxus wallichiana* complex (Taxaceae) based on herbarium material. *Botanical Journal of the Linnean Society 155*(3), 307-335. https://doi.org/10.1111/j.1095-8339.2007.00697.x

Omarova, P. (2018). Age intra population variability of generative bodies of *Taxus baccata* in foothill of Dagestan. *12*, 129-135. https://doi.org/10.17513/use.36916.

Paridari, I. C., Jalali, S. G., Sonboli, A., Zarafshar, M. & Bruschi, P. (2013). Leaf macro-and micro-morphological altitudinal variability of *Carpinus betulus* in the Hyrcanian forest (Iran). *Journal of Forestry Research, 24*, 301-307. https://doi.org/10.1007/s11676-013-0353-x

Pasini D. & Mirjalili V. W. (2006). The optimized shape of a leaf petiole transactions on *Ecology and the Environment, 87*, 35. https://doi.org/10.2495/dn060041

Perrin, P. & Mitchell, F. (2013). Effects of shade on growth, biomass allocation and leaf morphology in European yew (*Taxus baccata* L.). *European Journal of Forest Research, 132,* 211-218. https://doi.org/10.1007/s10342-012-0668-8

Pescador, D., Bello, F., Valladares, F. & Escudero, A. (2015). Plant trait variation along an altitudinal gradient in mediterranean high mountain grasslands: controlling the species turnover effect. *Plos One 10*(3), e0118876. https://doi.org/10.1371/journal.pone.0118876

Pfennigwerth, A., Bailey, J.K. & Schweitzer, J.A. (2017). Trait variation along elevation gradients in a dominant woody shrub is population-specific and driven by plasticity. https://doi.org/10.1093/aobpla/plx027

Read, Q.D., Moorhead, L.C., Swenson,N.G., Bailey,J.K. & Sander, N.J. (2014). Climate change and species range shifts convergent effects of elevation on functional leaf traits within and among species. *The Annals of Botany Plants, 9*(4), plx027, https://doi.org/10.1111/1365-2435.12162

Robakowski, P., Pers-Kamczyc, E., Ratajczak, E, Thomas P. A., Ye Zi-Piao, Rabska M. & Iszkuło, G. (2018). Photochemistry and antioxidative capacity of female and male *Taxus baccata* L. acclimated to different nutritional environments. *Frontiers in Plant Science, Plant Physiology, 9*, 742 https://doi.org/10.3389/fpls.2018.00742

Roche, P., Diaz Burlinson, N. & Gachet, S. (2004). Congruency analysis of species ranking based on leaf traits: Which traits are the more reliable*. Plant Ecology, 174,* 37-48. https://doi.org/10.1023/b:vege.0000046056.94523.57

Romo, A., Iszkuło, G., Seghir Taleb, M., Walas, Ł. & Boratyński, A. (2017). *Taxus baccata* in Morocco: a tree in regression in its southern extreme. *Dendrobiology,* 78, 63-74. https://doi.org/10.12657%2Fdenbio.078.007

Sarmoum, M., Navarro-Cerrilo, R. & Guibal, F. (2019). Bilan actuel et rétrospectif du dépérissement du cèdre de l'Atlas dans le Parc national de Theniet El Had (Algérie). *Bois et Forêts des Tropiques*, 342. https://doi.org/10.19182/bft2019.342.a31636

Sattarian, A., Akbarian, M., Zarafshar, M., Bruschi, P & Fayyaz, P. (2011). Phenotypic variation and leaf fluctuating asymmetry in natural populations of *Parrotia persica* (Hamamelidaceae), an endemic species from the Hyrcanian forest (Iran). *Acta Botanica Mexicana, 97,* 65-81.

Schirone, B., Caetano Ferreira, R., Vessella, F., Schirone, A. Piredda, R. & Cosimosi Simeone, M. (2010). *Taxus baccata* in the Azores: A relict form at risk of imminent exticinction. *Biodiversity and Conservation 19*(6), 1547-1565. https://doi.org/10.1007/s10531-010-9786-0

Shah, A., Li, D.Z., Möller,M., Gao,L., Hollingsworth, M.L. & Gibb, M. (2008). Delimitation of *Taxus fuana* Nan Li & R.R. Mill (Taxaceae) based on morphological and molecular data. *Taxon, 57*(1), 211-222. https://doi.org/10.2307/25065961

Spjut R. W. (2007 a). A phytogeographical analysis of *Taxus* (Taxaceae) based on leaf anatomical characters. *Journal of the Botanical Research Institute of Texas, 1*(1), 291-332.

Spjut R. W. (2007 b). Taxonomy and nomenclature of *Taxus* (Taxaceae). *Journal of the Botanical Research Institute of Texas, 1*(1),: 203-289.

Stefanović M., Nikolic, B., Matic, R., Popović, Z., Vidaković, V. & Bojovic, S. (2017). Exploration of sexual dimorphism of *Taxus baccata* L. needles in natural populations. *Trees, 31*, 1697–1710. https://doi.org/10.1007/s00468-017

Svenning J.C. & Magard E. (1999). Population ecology and conservation status of the last natural population of English yew *Taxus baccata* in Denmark. *Biological Conservation, 88*(2), 173-182. https://doi.org/10.1016/s0006-3207(98)00106-2

Thomas, P. A. & Polwart A. (2003). *Taxus baccata* L. *Journal of Ecology, 91,* 489-524. https://doi.org/10.1046/j.1365-2745.2003.00783.x

Tumpa, K., Liber, Z., Šatović, Z., Medak, J., Idžojtić, M., Vidaković, A., Vukelić, J.,Šapić, I., Nikl, P. & Poljak, I. (2022). High level of phenotypic differentiation of common yew (*Taxus baccata* L.) populations in the north-western part of the Balkan Peninsula. *Forests*, *13*, 78. https://doi.org/10.3390/f13010078

Vessella, F., Simeone, M.C., Fernandes, F.M., Schirone, A., Pires Gomes, M. & Schirone, B. (2013). Morphological and molecular data from Madeira support the persistence of an ancient lineage of *Taxus baccata* L. in Macaronesia and call for immediate conservation actions. *Caryologia, 66*, 162–177. https://doi.org/10.1080/00087114.2013.821842

Vivian, M., Robert, C., Godfree, M. J., Colloff, C., Ellery M. & Marshall, D. J.; (2014). Wetland plant growth under contrasting water regimes associated with river regulation and drought: implications for environmental water management. *Plant Ecology, 215*, 997–1011. http://doi.org/10.1007/s11258-014-0357-4.

Wyka, T., Robakowski, P. & Żytkowiak, R. (2007). Leaf age as a factor in anatomical and physiological acclimative responses of *Taxus baccata* L. needles to contrasting irradiance environments. *Photosynthesis Research, 95*(1), 87-99. http://doi.org/10.1007/s11120-007-9238-1

Yahi N. (2007). Les cédraies d'Algérie : phytoécologie, phytosociologie, dynamique et conservation des peuplements. Thèse Doctorat, UniversiteSTHB, Alger, 265p.

Zarek, M. (2007). Variability of morphological features of needles, shoots and seeds of *Taxus baccata* L. in nature reservations of southern Poland. *Electronic Journal of Polish Agricultural Universities*, *10*. https://doi.org/10.1007/s11270-020-04932-0

STUDIA UNIVERSITATIS BABEŞ-BOLYAI BIOLOGIA, 71, 1, 2026 (p. 125-159)
doi:10.24193/subbbiol.2026.1.07

Phylogenetic structuring of intrinsic disorder and charge patterning in vertebrate Ermin

Ilka Koszorus[1]✉ and Ferencz Kósa[1,2]

[1]*Babeş-Bolyai University, Hungarian Department of Biology and Ecology, Cluj-Napoca, Romania;*
[2]*Babeş-Bolyai University, Center for Systems Biology, Biodiversity and Bioresources, Sociobiology and Insect Ecology Lab, Cluj-Napoca, Romania;*
✉***Corresponding author, E-mail: koszorusilka@gmail.com.***

Article history: Received 27 February 2026; Revised 17 June 2026; Accepted 17 June 2026; Available online 25 June 2026

Abstract. Ermin (ERMN) is an oligodendrocyte-enriched cytoskeletal protein implicated in myelin sheath formation and stability. Although previously described as intrinsically disordered, its evolutionary diversification across vertebrates has not been systematically examined. Here, we performed a comprehensive phylogenetic and comparative analysis of 159 vertebrate Ermin orthologs to characterize the evolution of intrinsic disorder and physicochemical sequence properties. Across all major vertebrate clades, Ermin consistently fulfilled criteria for highly intrinsically disordered proteins based on Average Disorder Score (ADS) and Percent of Predicted Disordered Residues (PPDR), demonstrating that extensive disorder represents a deeply conserved structural feature. Despite this conserved disorder abundance, the architectural organization of intrinsically disordered regions (IDRs) varied among lineages. Mammalian Ermin proteins exhibited increased fragmentation of long IDRs compared to birds, which retained largely continuous long-IDR segments, indicating that IDR partitioning evolves independently of total disorder content. Amino-acid composition analyses revealed conserved depletion of aromatic and bulky hydrophobic residues across vertebrates, combined with clade-structured variation in charged and disorder-promoting residues.

Phylogenetic generalized least squares (PGLS) analyses detected strong phylogenetic signal across disorder and electrostatic traits. Root-to-tip regressions revealed a significant directional decrease in mean hydrophobicity across vertebrate divergence, whereas disorder propensity and sequence charge decoration (SCD) did not exhibit consistent monotonic trends. Fraction of charged residues (FCR) showed more modest evolutionary variation. Multivariate models further demonstrated that hydrophobicity, global charge density, and sequence charge decoration independently contribute to evolutionary variation in Ermin sequence architecture.

Together, these results indicate that Ermin evolution is characterized by architectural remodeling and compositional retuning of an already disordered scaffold rather than by progressive increases in intrinsic disorder. This study highlights IDR fragmentation and electrostatic organization as key evolutionary dimensions shaping the diversification of intrinsically disordered proteins across vertebrate lineages.

Keywords: Ermin (ERMN), Intrinsically disordered proteins (IDPs), Intrinsic disorder evolution, Sequence charge decoration (SCD), Phylogenetic comparative analysis (PGLS)

Introduction

Myelination is a defining feature of vertebrate nervous systems, enabling the rapid saltatory propagation of action potentials along axons and thereby ensuring efficient neural signalling (Nave and Werner, 2014). Proper myelin formation and maintenance rely on a limited set of highly specialized dosage-sensitive proteins that coordinate membrane organization with the underlying cytoskeleton (Boggs, 2006; Nawaz *et al.*, 2015). Several myelin-associated proteins deviate from classical globular architectures and are enriched in intrinsically disordered regions, a property linked to dynamic regulatory roles rather than static structural support (Jahn *et al.*, 2009; Harauz and Boggs, 2013). Despite their functional importance, comparative and evolutionary analyses of such proteins remain limited, particularly beyond mammalian systems (Baumann and Pham-Dinh, 2001; Aggarwal *et al.*, 2011; Nave and Werner, 2014). Notably, myelin is unusually enriched in proteins with extensive intrinsic disorder, many of which are involved in myelin regulation and are affected in demyelinating disease (Raasakka and Kursula, 2020).

Intrinsically disordered proteins (IDPs) and regions (IDRs) are widespread in eukaryotic proteomes and are particularly enriched among proteins involved in transcriptional regulation, signalling, and cytoskeleton-associated functions (Wright and Dyson, 1999; Uversky *et al.*, 2000; Bellay *et al.*, 2011; van der Lee *et al.*, 2014; Peng *et al.*, 2015; Holehouse and Kragelund, 2024). Large-scale analyses of disease-associated variants have shown that intrinsically disordered regions are frequent targets of pathogenic mutations, despite lacking stable tertiary structure. Approximately 20–25% of disease-related mutations map to predicted disordered regions, indicating that intrinsic disorder is neither evolutionarily neutral nor functionally dispensable (Vacic and Iakoucheva, 2012). Rather than disrupting global folding, disease-associated mutations in disordered regions often perturb short functional elements, including molecular recognition motifs, binding interfaces, or charge-patterned segments that are critical for interaction specificity and regulatory function (Vacic *et al.,* 2012). Together, these findings establish intrinsic disorder as a functionally constrained and disease-relevant sequence feature, motivating comparative analyses of disorder, amino-acid composition, and charge-related properties in proteins such as Ermin.

Ermin (also known as Juxtanodin, Fig. 1) was first described in 2005 as a previously uncharacterized protein enriched in the central nervous system (Zhang *et al.*, 2005). Early sequence analyses revealed the presence of a C-terminal actin-binding region with similarity to ezrin–radixin–moesin (ERM) family proteins, despite the absence of a canonical ERM FERM domain (Zhang *et al.*, 2005; Brockschnieder *et al.,* 2006). Subsequent experimental studies demonstrated that Ermin associates with F-actin and modulates cytoskeleton-dependent cell morphology, including effects on cellular process formation and arborization (Zhang *et al.*, 2005; Meng *et al.*, 2010; Wang *et al.*, 2011; Ruskamo *et al.*, 2012). Although Ermin is enriched in oligodendrocytes and localizes to non-compact myelin regions during late stages of myelination, its precise mechanistic role in myelin formation and maintenance remains unclear (Brockschnieder *et al.*, 2006; Wang *et al.*, 2020). Importantly, Ermin expression is not restricted to oligodendrocytes, as it is also detected in retinal pigment epithelial cells, where it localizes to actin-rich subcellular regions and influences actin organization (Liang *et al.*, 2018). Taken together, existing evidence supports viewing Ermin primarily as an actin-associated cytoskeletal regulator with cell-type-specific functional consequences rather than as a structural myelin protein per se, despite its enrichment in oligodendrocytes (Zhang *et al.*, 2005; Brockschnieder *et al.*, 2006; Wang *et al.*, 2020).

Altered ERMN expression or genetic variation has been reported in multiple neurological and neurodevelopmental disorders, including autism spectrum disorder, epilepsy, schizophrenia, and multiple sclerosis (Salek Esfahani *et al.*,

2019; Wang *et al.*, 2020; Shiva *et al.*, 2021, Ziaei *et al.*, 2022). The ERMN gene is located within the AUTS5 locus, a genomic region that has been repeatedly implicated in autism, language impairment, and cognitive traits (Shiva *et al.*, 2021). Both reduced expression and dosage imbalance of ERMN have been associated with disease phenotypes, suggesting that tight regulation of Ermin sequence and expression is biologically important (Homs *et al.*, 2016; Shiva *et al.*, 2021). Together, these disease associations provide strong motivation for investigating Ermin's sequence properties and evolutionary constraints.

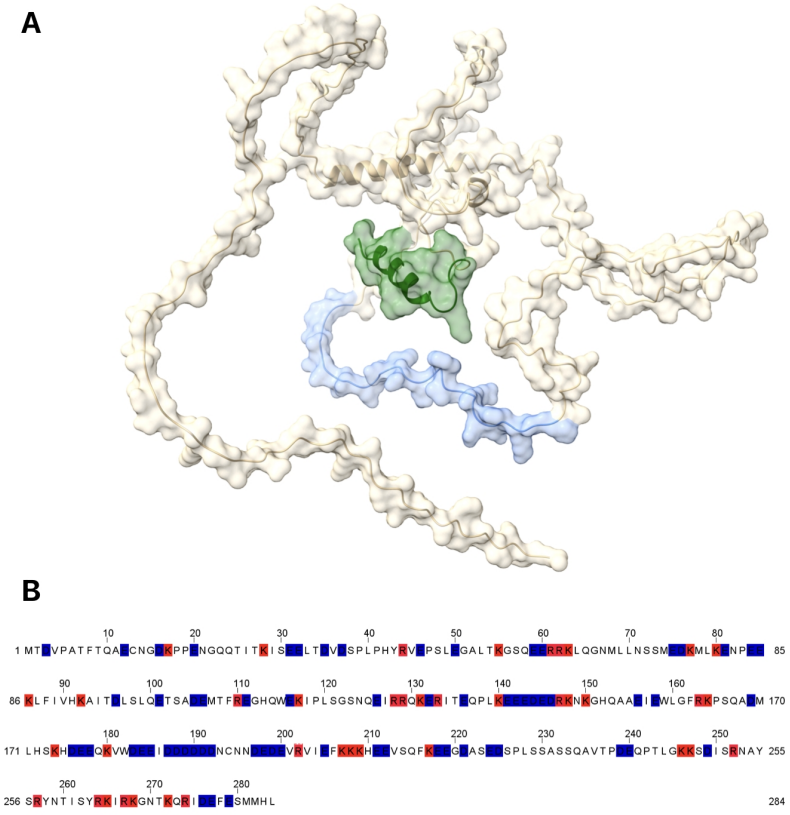


Figure 1. Structural and functional features of human Ermin (UniProt ID: Q8TAM6). AlphaFold-predicted three-dimensional structure of Ermin (AF-Q8TAM6-F1-v6; https://alphafold.ebi.ac.uk/entry/AF-Q8TAM6-F1) visualized using ChimeraX (Pettersen *et al.*, 2021). (A) The experimentally characterized F-actin binding site (Zhang *et al.*, 2005; Ruskamo *et al.*, 2012) is highlighted in green, and the acidic region (Ruskamo *et al.*, 2012) is shown in blue. (B) Amino acid sequence of human Ermin (UniProt ID: Q8TAM6), with positively charged residues (Arg, Lys) indicated in red and negatively charged residues (Asp, Glu) in blue.

Ermin lacks extensive stable secondary or tertiary structure and is predicted to contain large IDRs, consistent with sequence-based analyses and functional studies of the protein (Ruskamo *et al*., 2012). Intrinsically disordered regions are well suited to signaling, scaffolding, cytoskeletal regulation, and sensitivity to amino-acid composition and charge patterning, features that are particularly relevant for cytoskeleton-associated proteins (Wright and Dyson, 1999; Uversky *et al.*, 2000).

Despite growing interest in Ermin's cellular and pathological relevance, its evolutionary history across vertebrates has not been systematically examined. It remains unclear how conserved Ermin sequence features are across major vertebrate clades, whether intrinsic disorder and charge-related properties are evolutionarily stable or exhibit clade-specific variation, and how amino-acid compositional biases within disordered regions differ among lineages. Moreover, most existing studies of Ermin are restricted to mammalian systems, which limits broader evolutionary inference.

The primary objective of this study was to investigate the evolutionary diversification of intrinsic disorder in vertebrate Ermin using a comprehensive phylogenetic and sequence-based framework. Specifically, we aimed to (i) determine whether Ermin is consistently classified as a highly intrinsically disordered protein across major vertebrate lineages; (ii) characterize the architectural organization of intrinsically disordered regions (IDRs), with particular emphasis on the distribution, continuity, and fragmentation of long IDRs; (iii) quantify lineage-specific variation in amino-acid composition, hydrophobicity, and electrostatic features, including fraction of charged residues (FCR), net charge per residue (NCPR), and sequence charge decoration (SCD); and (iv) assess evolutionary trends and phylogenetic structuring of these properties using phylogenetically informed comparative methods.

By integrating disorder prediction, polymer-inspired electrostatic metrics, and phylogenetic generalized least squares (PGLS) analyses, we sought to test whether Ermin evolution is characterized by monotonic changes in disorder content or instead by lineage-specific remodelling of disorder architecture and sequence composition. Through this approach, we aim to establish a quantitative framework for understanding how intrinsically disordered proteins diversify over deep evolutionary timescales while maintaining conserved functional disorder.

Materials and methods

Sequence retrieval

Ermin amino acid sequences were collected using the same NCBI BLAST-based workflow (https://blast.ncbi.nlm.nih.gov) described in our previously published myelin basic protein study (Koszorus and Kósa, 2025). Sequence collection was restricted to species with available, well-annotated whole-genome assemblies. BLASTp searches were conducted separately for each major vertebrate clade using clade-specific Ermin reference sequences.

For cartilaginous fishes (*Chondrichthyes*), *Callorhinchus milii* (XP_007888107.2) was used as the reference sequence. In ray-finned fishes (*Actinopterygii*), two reference sequences were employed to account for the pronounced length variability observed among Ermin homologs in this clade: *Danio rerio* (XP_073761994.1), and *Perca fluviatilis* (XP_039674551.1). Single representative reference sequences were used for amphibians (*Xenopus tropicalis*, XP_002933329.2), reptiles (*Gekko japonicus*, XP_015267048.1), birds (*Gallus gallus*, XP_040531879.1), and mammals (*Homo sapiens*, Q8TAM6).

All BLAST parameters, filtering criteria, and orthology assignment steps were identical to those described previously (Koszorus and Kósa, 2025). In total, 159 Ermin sequences were curated, with clade-specific distribution (cartilaginous fishes: 13, ray-finned fishes: 50, amphibians: 16, reptiles: 17, birds: 19, and mammals: 44).

Prediction of protein disorder

The intrinsic disorder of Ermin proteins was predicted using the Rapid Intrinsic Disorder Analysis Online (RIDAO) platform (https://ridao.app, Dayhoff and Uversky, 2022), which incorporates six per residue predictors of disorder: PONDR VLXT (Li *et al.*, 1999), PONDR VL3 (Radivojac *et al.*, 2003), PONDR VSL2B (Peng *et al.*, 2006), PONDR FIT (Xue *et al.*, 2010), IUPred (short and long modes) (Dosztányi *et al.*, 2005, Erdős and Dosztányi, 2020). All predictors provide a score that characterizes the disorder propensity of each position along the sequence. Residues with scores above 0.5 are predicted as disordered, while scores below 0.5 indicate order. Predictor outputs were aggregated and exported as mean disorder profiles (MDP). The mean disorder profile was used to identify the intrinsically disordered regions (IDRs) in the protein, and allowed us to calculate the disorder properties related to the whole protein, such as the average disorder score (ADS), percentages of predicted disordered residues (PPDR), and the length and positions of the individual disordered regions.

Identification, classification, and evolutionary comparison of Ermin IDR architecture

Residues were classified as disordered (IDR) or ordered using a disorder threshold of 0.5, such that residues with disorder scores ≥0.5 were assigned to IDRs, while residues below this threshold were considered ordered. According to the length, intrinsically disordered regions were classified in short IDRs (SDR, 5-29 consecutive residues with disorder scores ≥0.5) (Monzon *et al.*, 2020) and long IDRs (LDR, ≥ 30 residues) (van der Lee *et al.*, 2014; Monzon *et al.*, 2020).

Each IDR was further assigned a positional class (N-terminal, internal, or C-terminal) based on its location relative to the full ungapped protein sequence. For each Ermin ortholog, summary descriptors were calculated, including the number of long and short IDRs, their lengths, and the fraction of amino acids residing in each IDR class.

To visualize the phylogenetic distribution of IDR architecture, IDR maps were plotted alongside a Neighbour-Joining phylogenetic tree inferred from aligned Ermin sequences (Fig. 2). IDRs were shown as horizontal bars scaled to sequence length, enabling direct comparison of disorder organization across clades.

Quantification of IDR fragmentation

To decouple overall disorder abundance from architectural organization, new IDR fragmentation indices were defined. Fragmentation indices normalize the number of disordered segments by the fraction of disordered amino acids, thereby capturing how disorder is partitioned rather than how much disorder is present. Two indices were calculated per sequence: Total-IDR fragmentation index (FI_{total}), $FI_{total} = N_{total\ IDRs} / (fraction\ of\ amino\ acids\ in\ long + short\ IDRs)$, where $N_{total\ IDRs}$ is the total number of IDRs; and Long-IDR fragmentation index (FI_{long}), $FI_{long} = N_{long\ IDRs} / fraction\ of\ amino\ acids\ in\ long\ IDRs$, where $N_{long\ IDRs}$ is the number of long IDRs. Fragmentation indices were summarized by major vertebrate clades and compared statistically using non-parametric tests.

Amino-acid compositional bias of intrinsically disordered regions

To further characterize intrinsically disordered regions, we compared the amino-acid composition of residues classified as disordered with that of residues located in ordered regions. Amino-acid compositional bias between intrinsically disordered regions (IDRs) and ordered regions was quantified using per-residue disorder predictions and protein sequence data (Vacic *et al.*, 2007, Djulbegovic and Uversky, 2022).

For each amino acid, pooled frequencies were calculated separately for IDR and ordered regions within the same protein sequences, and aggregated across all sequences. Relative amino-acid composition was then computed as *Relative composition* = $(f_{IDR} - f_{ordered})/f_{ordered}$, where f_{IDR} and $f_{ordered}$ denote the fractional abundance of a given amino acid in IDR and ordered regions, respectively. Positive values indicate enrichment in IDRs, whereas negative values indicate depletion relative to ordered regions. Confidence intervals for relative composition estimates were obtained by non-parametric bootstrap resampling of protein sequences (10,000 iterations). In each bootstrap replicate, sequences were sampled with replacement, residue counts were pooled, and relative composition values were recalculated. The 2.5th and 97.5th percentiles of the bootstrap distributions were taken as the 95% confidence intervals. Amino acids whose confidence intervals overlapped zero were considered not significantly enriched or depleted.

Phylogenetic comparative framework

To investigate the evolutionary dynamics of Ermin sequence properties across vertebrates, we employed phylogenetic comparative analyses that explicitly account for non-independence among species due to shared evolutionary history. Phylogenetic generalized least squares (PGLS) regression was used as the primary statistical framework, following the classical formulation of phylogenetic regression and its subsequent extensions for comparative data analysis (Grafen, 1989; Martins and Hansen, 1997; Freckleton *et al.*, 2002). This approach incorporates the expected covariance structure among species derived from a phylogenetic tree and allows hypothesis testing on trait evolution while controlling for phylogenetic relatedness.

Phylogenetic tree reconstruction

A species phylogeny was constructed using amino-acid sequences aligned with ClustalW and inferred using the Neighbor-Joining (NJ) method implemented in MEGA12. The NJ algorithm provides an efficient distance-based approximation of phylogenetic relationships and is widely used for comparative and exploratory evolutionary analyses when the primary goal is to capture relative divergence among taxa rather than time-calibrated histories (Saitou and Nei, 1987; Kumar *et al.*, 2018). Branch lengths correspond to sequence divergence. For PGLS analyses, the exact topology and relative branch length structure are more critical than time calibration, and distance-based Neighbor-Joining trees are sufficient for estimating phylogenetic covariance among taxa when the objective is to control for shared ancestry rather than to reconstruct time-calibrated evolutionary histories (Grafen, 1989; Freckleton *et al.*, 2002).

Prior to PGLS analyses, the tree was midpoint-rooted, an established rooting approach when a reliable outgroup is unavailable, which places the root at the midpoint of the longest path between taxa and preserves relative branch length structure (Farris, 1972; Hess and De Moraes Russo, 2007). Because branch lengths represent evolutionary divergence rather than absolute time, root-to-tip distances were interpreted as relative evolutionary distances along the phylogeny rather than chronological age, and were used to assess directional trends in trait evolution rather than absolute evolutionary rates. Phylogenetic signal was quantified using Pagel's λ, which rescales internal branch lengths to best fit the observed trait covariance under a Brownian motion model (Pagel, 1999). λ was estimated by maximum likelihood for each fitted model and interpreted as a measure of the degree to which trait variation reflects phylogenetic structure, with values approaching 1 indicating strong phylogenetic signal and values near 0 indicating phylogenetic independence (Blomberg *et al.*, 2003).

Sequence-derived trait calculation

For each Ermin sequence, multiple physicochemical and disorder-related traits were computed. Disorder scores were obtained from per-residue disorder predictions (RIDAO MDP) using the mean disorder probability (MDP) score across residues.

Fraction of charged residues (FCR) was defined as the fraction of positively and negatively charged residues (FCR = $f^+ + f^-$, where $f^+ = N^+/N$ and $f^- = N^-/N$, N is total of residues in the sequence, and $N^+$and N^- the number of positive, negatively charged residues) (Das and Pappu, 2013; Das *et al*., 2015). Net charge per residue (NCPR) was calculated as the difference between positively and negatively charged residues (NCPR = $f^+ – f^–$), preserving the sign of the net charge (Das and Pappu, 2013; Das *et al*., 2015). Sequence charge decoration (SCD) was computed following established definitions that quantify charge patterning along the sequence, with higher values indicating greater segregation of like charges (Sawle and Ghosh, 2015; Ghosh *et al*., 2022). The mean hydrophobicity was calculated using the Kyte–Doolittle hydrophobicity scale (Kyte and Doolittle, 1982).

Clade-wise PGLS analyses and pairwise contrasts

Univariate PGLS models were fitted separately for each sequence property to test for differences among major vertebrate clades (*Chondrichthyes, Actinopterygii, Amphibia, Reptilia, Aves,* and *Mammalia*), with clade identity treated as a categorical predictor. To formally assess whether clade identity improved model fit beyond phylogenetic structure alone, each full model

(including clade identity) was compared to a reduced, intercept-only model using likelihood-ratio tests (Burnham and Anderson, 2002). The resulting likelihood-ratio statistics and associated p-values were used to evaluate whether inclusion of clade identity improved model fit relative to phylogenetic covariance alone. Pairwise contrasts between clade means were derived from the fitted PGLS models to identify which clade pairs contributed most strongly to overall differences. To control for multiple testing across contrasts, p-values were adjusted using the Holm–Bonferroni procedure, which controls the family-wise error rate (Holm, 1979).

Root-to-tip trend analyses

To test for directional evolutionary trends, additional PGLS models were fitted in which each trait was regressed against root-to-tip phylogenetic distance, calculated as the sum of branch lengths from the root to each species. Under this framework, a significant regression slope, if detected, indicates a directional trend in trait evolution, whereas non-significant slopes are consistent with clade-structured or stationary evolution (Hunt, 2006). Root-to-tip analyses were conducted within a Brownian/PGLS framework rather than Ornstein–Uhlenbeck models to avoid over-parameterization and misinterpretation in the presence of strong phylogenetic signal (Cooper *et al.*, 2016).

Phylogenetic multiple regression

To assess the independent contributions of multiple sequence properties, a phylogenetic multiple regression was fitted including sequence length, mean hydrophobicity, FCR, NCPR, and SCD as predictors. Regression coefficients were standardized to allow direct comparison of effect sizes under the phylogenetic covariance structure. This analysis complements univariate and clade-wise comparisons by identifying which properties explain evolutionary variation in Ermin sequence architecture when other correlated traits are held constant.

Results

Vertebrate Ermin Proteins Are Highly Disordered

Proteins can be classified into structural disorder categories based on the Average Disorder Score (ADS) and the Percentage of Predicted Disordered Residues (PPDR), typically as highly ordered, moderately disordered, or highly disordered (Gsponer *et al.* 2008; Rajagopalan *et al.* 2011; Djulbegovic and Uversky, 2022). In our dataset, all 159 analyzed vertebrate Ermin sequences exhibited ADS values above 0.5 and PPDR values exceeding 30% (Table 1.).

According to commonly accepted classification criteria, these parameters place Ermin in the highly disordered category. This classification is consistent with previous observations for Myelin Basic Protein, which similarly falls within the highly disordered class (Koszorus and Kósa, 2025).

Table 1. Intrinsic disorder metrics and IDR architectural parameters of vertebrate Ermin across major clades. Descriptive summary of intrinsic disorder–related properties for 159 vertebrate Ermin sequences grouped by major lineages. Parameters include Average Disorder Score (ADS), Percent of Predicted Disordered Residues (PPDR), total number of intrinsically disordered regions (IDRs), number and mean length of long (LDRs; ≥30 residues) and short (SDRs; 5–29 residues) disordered regions, mean number of LDRs and SDRs per sequence, and the percentage of residues within LDRs and SDRs. Values are reported as clade-wise means with observed minimum–maximum ranges and are presented for descriptive purposes only.

Parameter	*Chondrichthyes*	*Actinopterygii*	*Amphibia*	*Reptilia*	*Aves*	*Mammalia*
Mean sequence length (aa)	419	268.2	245.25	255.76	252.42	279.14
PPDR (%)	92.7	88.19	87.67	92.93	97.46	86.45
ADS	0.72	0.72	0.713	0.724	0.81	0.69
Nr of IDRs/sequence (min-max)	1-6	1-6	1-4	1-3	1-2	2-4
Mean nr of IDRs/sequence	2.92	2.34	2.25	1.82	1.53	2.77
Percent of residues in IDR (%)	91.03	86.23	87.59	90.03	96.97	85.80
Nr of SDRs/sequence (min-max)	0-2	0-4	0-1	0-2	0-1	0-1
Mean nr of SDRs/sequence	0.61	0.96	0.81	0.65	0.32	0.09
Mean length of SDR (aa)	9.33	17.18	16.61	26.77	23.66	17
Percent of residues in SDR (%)	1.41	5.34	5.53	6.77	2.93	0.55
Nr of LDRs/sequence (min-max)	1-4	0-3	1-3	1-2	1-2	2-3
Mean nr of LDRs/sequence	2.30	1.38	1.44	1.18	1.21	2.68
Mean length of LDR (aa)	206.57	175.72	161.80	193.11	211.39	92.20
Percent of residues in LDR (%)	89.61	80.89	82.05	83.26	94.04	85.25

Conserved long-IDR–rich architecture across vertebrate Ermin proteins

Across all examined vertebrate Ermin orthologs (n = 159), protein architecture was dominated by long intrinsically disordered regions. In every major clade, long IDRs accounted for the majority of disordered residues, indicating that extensive disorder is a deeply conserved feature of Ermin.

Phylogeny-aligned visualization of IDR architecture revealed that this conservation masks substantial variation in how long disordered residues are organized within individual proteins (Fig. 2). While overall disorder content remained high across clades (Table 1), the number, continuity, and positional arrangement of IDRs differed markedly.

Clade-specific organization and localization of IDRs

C-terminal long IDRs were nearly universal across tetrapods, suggesting strong functional constraint on this region. In contrast, internal long IDRs displayed pronounced lineage specificity: they were frequent in mammals but rare or absent in birds and reptiles (Fig. 2).

Short IDRs were common in actinopterygians and amphibians, often located at N-terminal or internal positions, but were strongly reduced in birds and mammals. Notably, mammalian Ermin sequences exhibited an almost complete absence of short IDRs despite maintaining extensive long-IDR coverage, indicating a shift in how disorder is partitioned rather than a reduction in overall disorder per se.

Mammalian Ermin shows increased fragmentation of long IDRs

Quantitative analysis of fragmentation indices revealed clear clade-specific differences. Birds (*Aves*) exhibited the lowest total-IDR fragmentation index (FI_{total}), with a narrow distribution centered below the global median (Fig. 3A). In contrast, mammals showed significantly elevated FI_{total} values relative to birds, reflecting increased segmentation of disordered segments.

This pattern was even more pronounced for long-IDR fragmentation (FI_{long}). Birds exhibited the lowest FI_{long} values, consistent with largely continuous long IDRs, whereas mammals displayed a strong and statistically significant increase in FI_{long} (Fig. 3B). Intermediate fragmentation levels were observed in actinopterygians, amphibians, and reptiles, while chondrichthyans showed high variance, indicating heterogeneous IDR architectures.

These results demonstrate that mammalian Ermin proteins are characterized by increased partitioning of long disordered regions, rather than by an increase in total disorder content.

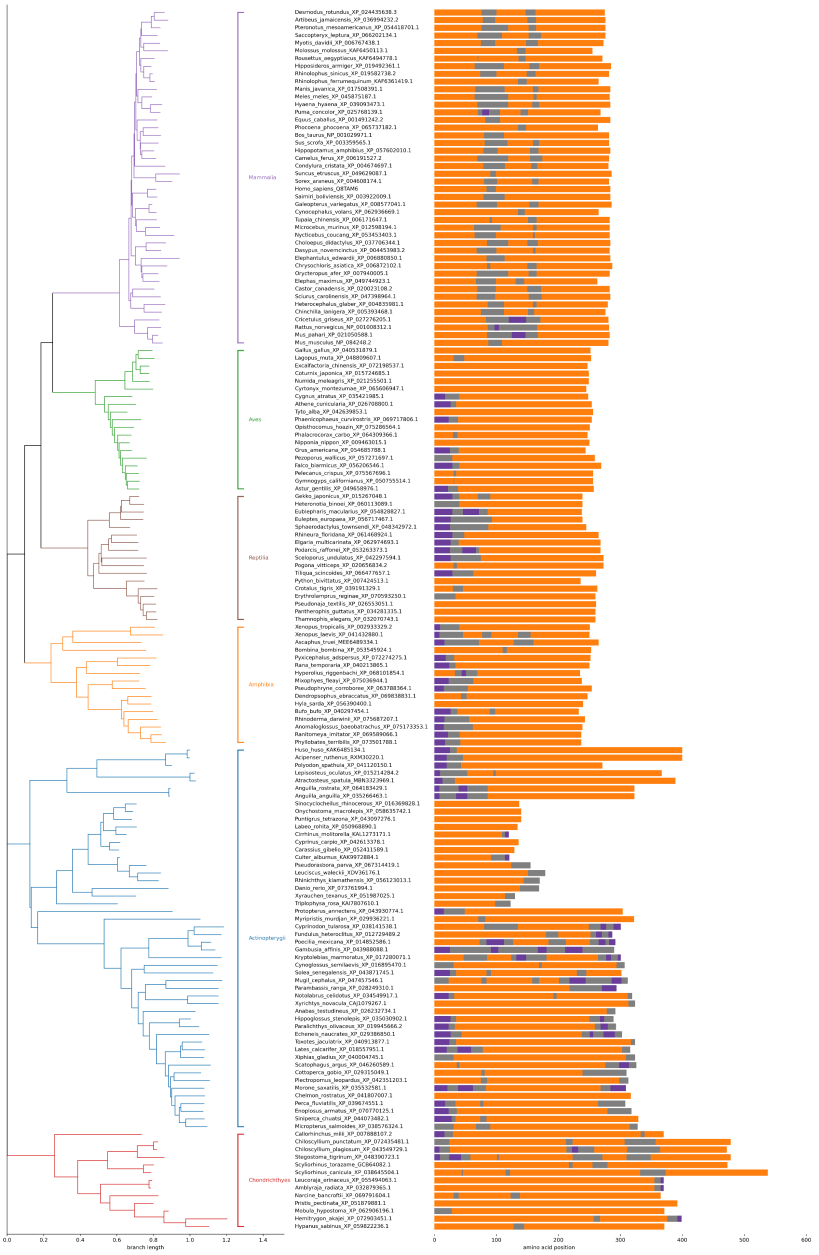


Figure 2. Phylogeny-linked IDR architecture of vertebrate Ermin. A phylogenetic tree of 159 Ermin sequences is shown with branches colored by major vertebrate clade. Horizontal bars depict per-sequence disorder architecture in ungapped sequence coordinates: ordered regions (grey), long IDRs (orange), and short IDRs (purple). Bars are aligned at the N-terminus and scaled by each sequence's ungapped length.

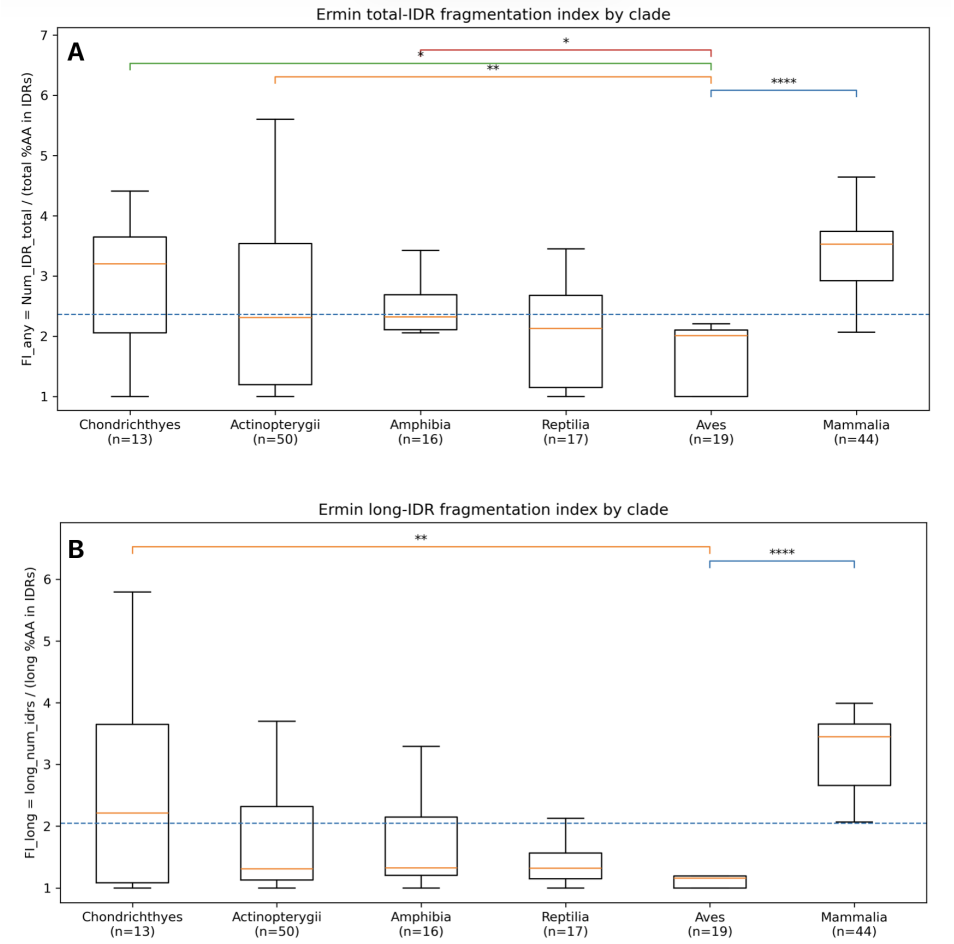


Figure 3. Clade-specific fragmentation of Ermin intrinsically disordered regions. Boxplots show the distribution of intrinsically disordered region (IDR) fragmentation indices across major vertebrate clades. (A) Total-IDR fragmentation index (FI_{total}), defined as the number of all IDRs normalized by the total fraction of amino acids classified as disordered. (B) Long-IDR fragmentation index (FI_{long}), defined as the number of long IDRs normalized by the fraction of amino acids classified as long-IDR. Boxes indicate the interquartile range with medians shown as horizontal lines; whiskers extend to 1.5× IQR. Numbers in parentheses denote the number of sequences per clade. The blue dashed line marks the global median fragmentation index across all Ermin sequences, serving as a vertebrate-wide reference. Statistical significance was assessed using Kruskal–Wallis tests followed by Dunn's post-hoc comparisons with Holm correction. Brackets indicate pairwise comparisons relative to the avian clade (***Aves***), which exhibits the lowest fragmentation and serves as a baseline for continuous IDR architecture (* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$).

Global features of vertebrate IDR composition

Analysis of amino-acid composition revealed pronounced differences between intrinsically disordered regions (IDRs) and ordered regions of Ermin proteins (Fig. 4). IDRs were significantly depleted in aromatic (tryptophan, tyrosine, phenylalanine, histidine) and large hydrophobic residues (isoleucine, leucine, valine, methionine), consistent with reduced hydrophobic packing requirements in disordered conformations. In contrast, IDRs exhibited strong enrichment in proline, as well as increased representation of polar and charged residues, including serine, glutamine, arginine, aspartate, and glutamate.

Proline showed the strongest and most consistent enrichment among all amino acids, reflecting its well-established role in disrupting secondary structure and promoting conformational flexibility. Charged residues displayed moderate but significant enrichment, supporting an IDR architecture dominated by electrostatic interactions rather than hydrophobic collapse. Cysteine was markedly depleted in IDRs, consistent with its preferential involvement in structured domains and disulfide-bond formation.

Bootstrap confidence interval analysis indicated that most observed compositional biases were statistically robust across sequences. Only a small subset of residues exhibited confidence intervals overlapping zero, indicating limited or variable enrichment across species. Across vertebrate clades, IDRs display both conserved disorder signatures and striking lineage-specific compositional specializations. Chondrichthyans exhibit predominantly polar, uncharged IDRs enriched in serine and threonine, with strong depletion of aromatic and hydrophobic residues and minimal charge enrichment, yielding a low-aromatic, low-hydrophobic architecture (Fig. S1A). In contrast, *Actinopterygii* show the most pronounced acidic signature, with strong enrichment of glutamate and aspartate and exceptional relative enrichment of proline, indicating a highly charged, electrostatically driven IDR composition (Fig. S1B). Amphibian IDRs retain acidic dominance but display increased compositional heterogeneity, with marked enrichment of asparagine and glutamate alongside variable contributions from basic and aromatic residues (Fig. S1C). Reptilian IDRs are distinguished by extreme and heterogeneous enrichment of histidine, threonine, and lysine, superimposed on conserved depletion of cysteine and bulky hydrophobics, suggesting clade-specific modulation of charge and potential pH sensitivity (Fig. S1D).

Aves exhibit the strongest low-complexity profile, characterized by enrichment of methionine, leucine, cysteine, proline, threonine, arginine and glutamate and profound depletion of branched hydrophobic residues (isoleucine, valine), as well as near absence of aromatics, reflecting highly flexible, compositionally simplified

IDRs (Fig. S1E). Finally, Mammalian IDRs combine strong enrichment of disorder-promoting residues — particularly arginine, and proline — with marked depletion of bulky aromatics, yielding a composition that integrates charge enrichment with maintained proline-driven structural disruption (Fig. S1F). Collectively, these patterns indicate progressive diversification of IDR electrostatic and low-complexity features across vertebrate evolution, with distinct clade-specific biases superimposed on a conserved depletion of aromatic and hydrophobic residues.

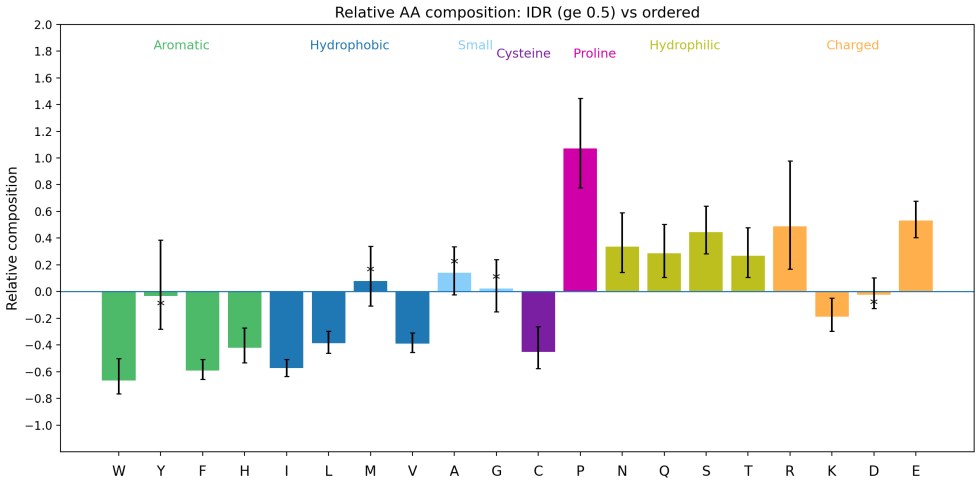


Figure 4. Relative amino-acid composition of intrinsically disordered regions (IDRs) versus ordered regions in Ermin proteins. Relative composition was calculated as $(f_{IDR} - f_{ordered}) / f_{ordered}$; positive values indicate enrichment in IDRs and negative values indicate depletion. Amino acids are grouped by physicochemical properties, with cysteine and proline shown separately. Error bars represent 95% bootstrap confidence intervals (10,000 sequence resamplings). Asterisks denote residues whose confidence intervals overlap zero (non-significant enrichment or depletion).

Phylogenetic distribution of Ermin sequence properties

Mapping z-score–normalized sequence properties onto the Ermin phylogeny revealed pronounced structuring across disorder, hydrophobicity, and electrostatic traits (Fig. 5). Closely related species clustered together in trait space, and distinct clade-level signatures were evident for multiple properties. Disorder propensity showed a non-monotonic distribution across vertebrates: early-diverging lineages, particularly chondrichthyans, already exhibited high disorder values; birds displayed the highest disorder overall; and mammals showed

comparatively lower disorder than birds and several fish lineages. Thus, high intrinsic disorder is not restricted to derived vertebrate groups and does not increase monotonically across the phylogeny.

In contrast, mean Kyte–Doolittle hydrophobicity showed a clearer lineage-dependent pattern, with mammals and reptiles exhibiting more negative values than most fish lineages, consistent with reduced hydrophobicity. Electrostatic properties also displayed strong lineage dependence: fraction of charged residues (FCR), net charge per residue (NCPR), and sequence charge decoration (SCD) each showed clade-specific patterns, with SCD in particular separating vertebrate groups in the phylogeny-aligned heatmap. Pagel's λ values were consistently high across traits ($\lambda \approx 0.85–1.0$), indicating substantial phylogenetic signal and strong covariance among species.

Clade-wise distributions and pairwise PGLS contrasts

Clade-wise comparisons of raw trait distributions confirmed the lineage-specific patterns observed in the phylogenetic mapping (Fig. 6). Disorder propensity exhibited substantial variation across clades but did not follow a simple monotonic trajectory. Birds exhibited the highest disorder distributions, whereas mammals showed lower disorder relative to birds and some early-diverging lineages. Mean hydrophobicity differed among clades, with mammals and reptiles tending toward more negative values, consistent with reduced hydrophobicity. Electrostatic traits also varied among clades. FCR and SCD showed broader interclade separation than NCPR, which remained within a relatively narrow range across vertebrates.

To formally evaluate clade effects while accounting for phylogenetic non-independence, univariate PGLS models including clade identity were compared with reduced intercept-only models using likelihood-ratio tests. In all cases, inclusion of clade identity did not significantly improve model fit after accounting for phylogenetic covariance (Table S2). These results indicate that the apparent clade-level differences observed in raw trait distributions are largely explained by shared evolutionary history rather than by statistically discrete shifts among categorical clades.

Consistent with this finding, pairwise PGLS contrasts among clade means revealed directional differences in several traits; however, none of the individual pairwise comparisons remained significant after Holm–Bonferroni correction for multiple testing (Table S1). The absence of statistically significant pairwise contrasts reflects the combined effects of strong phylogenetic signal (high Pagel's λ values), moderate within-clade variance, and conservative multiple-testing correction.

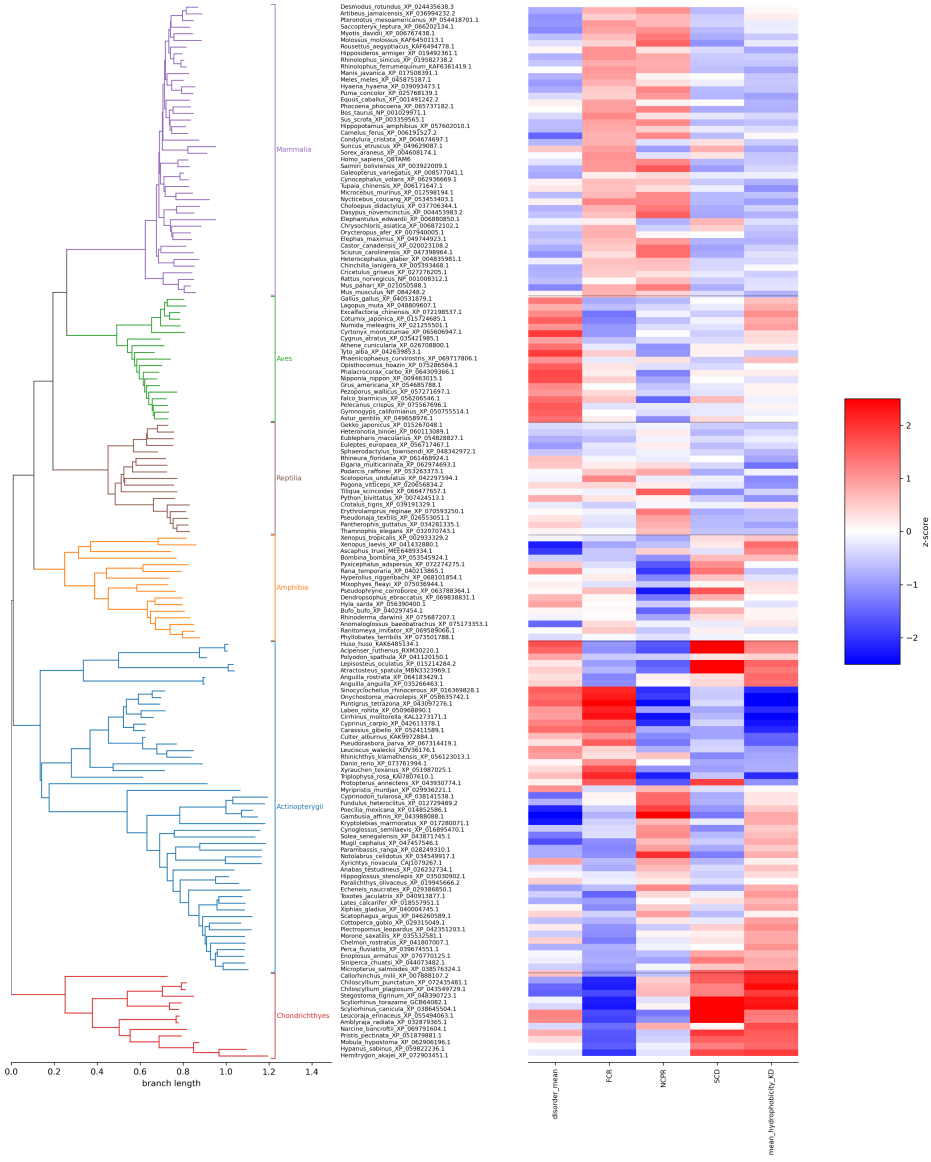


Figure 5. Phylogenetic distribution of disorder, mean hydrophobicity and electrostatic properties of Ermin across vertebrates. A species phylogeny of vertebrate Ermin sequences is shown on the left, with branches colored by major clades (*Chondrichthyes, Actinopterygii, Amphibia, Reptilia, Aves, Mammalia*). Brackets indicate clade boundaries. To the right, a heatmap displays z-score–normalized values of mean disorder, fraction of charged residues (FCR), net charge per residue (NCPR), sequence charge decoration (SCD) and mean hydrophobicity for each species, aligned to the corresponding tree tip. Red and blue indicate values above and below the overall mean, respectively.

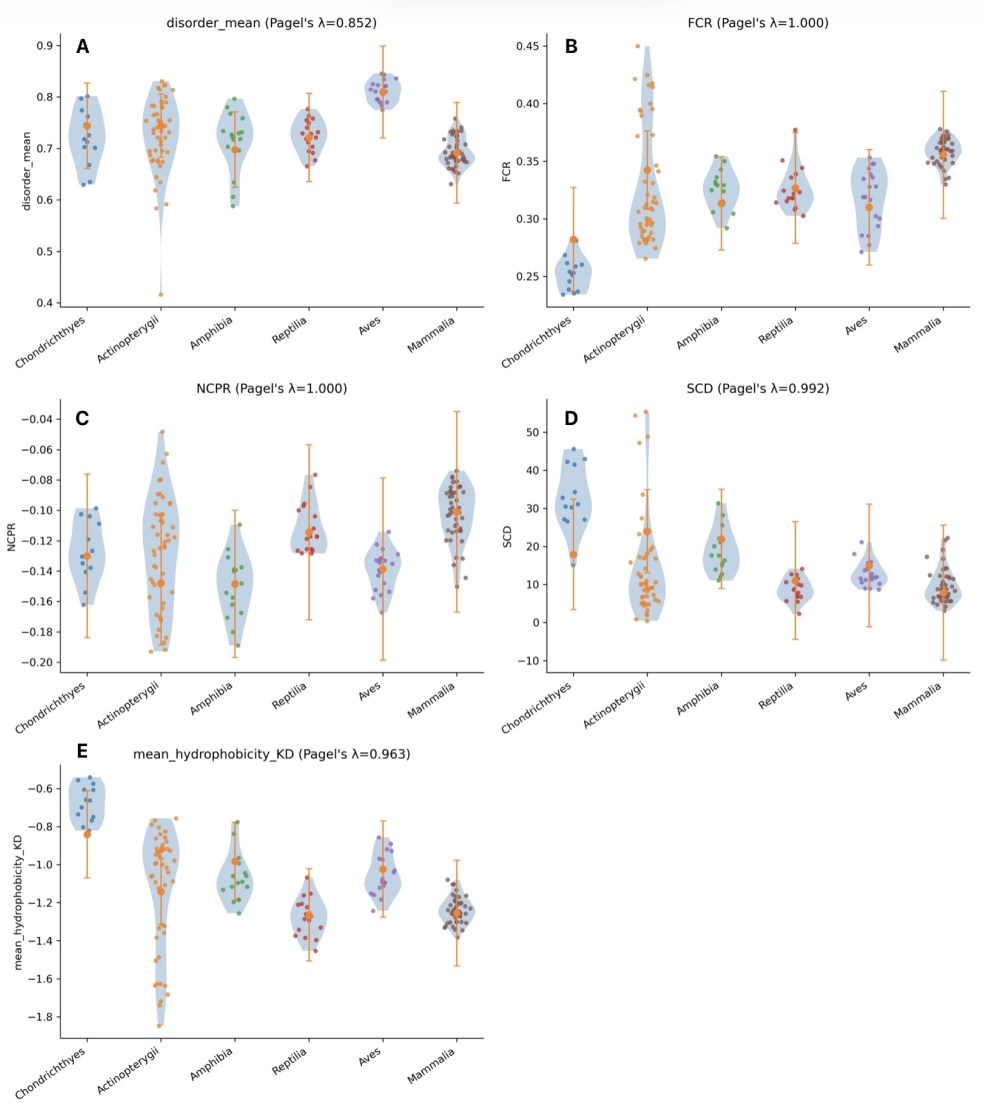


Figure 6. Evolutionary changes in intrinsic disorder, mean hydrophobicity and electrostatic properties of vertebrate Ermin. Violin plots show the distribution of sequence-level properties across major vertebrate clades (*Chondrichthyes*, *Actinopterygii*, *Amphibia*, *Reptilia*, *Aves*, *Mammalia*). Points represent individual species; violins indicate kernel density estimates. Orange circles and error bars denote phylogenetically corrected clade means ± 95% confidence intervals estimated using PGLS. (A) Mean disorder score, (B) fraction of charged residues (FCR), (C) net charge per residue (NCPR), (D) sequence charge decoration (SCD), (E) mean hydrophobicity. Pagel's λ values are shown for each trait, indicating strong phylogenetic signal, particularly for electrostatic properties.

Taken together, these analyses suggest that variation in ermin sequence properties is structured along the phylogeny in a gradual, lineage-dependent manner rather than partitioned into sharply distinct clade-specific regimes. The data therefore support a model of distributed evolutionary tuning across vertebrate lineages rather than discrete categorical shifts among major clades.

Root-to-tip analyses reveal limited directional trends

Root-to-tip PGLS regressions were performed to assess directional evolutionary trends across the vertebrate Ermin phylogeny (Table 2). Mean Kyte–Doolittle hydrophobicity exhibited a highly significant positive slope ($p = 0.0008$), indicating a progressive decrease in hydrophobicity (shift toward more negative KD values) with increasing phylogenetic distance from the root. Fraction of charged residues (FCR) also showed a statistically significant but modest negative slope ($p = 0.0152$), suggesting a weak directional reduction in overall charge density. In contrast, net charge per residue (NCPR), sequence charge decoration (SCD), and mean disorder propensity did not display significant monotonic trends. Pagel's λ values were high for all traits, indicating substantial phylogenetic signal.

Collectively, these findings indicate that while hydrophobicity — and to a lesser extent overall charge density — exhibits directional evolutionary change, variation in disorder propensity and charge patterning is better explained by lineage-specific restructuring rather than by uniform trends across the vertebrate phylogeny.

Multivariate determinants of Ermin sequence evolution

To assess the independent contributions of multiple sequence features, a multivariate PGLS regression was fitted including sequence length, mean hydrophobicity, FCR, NCPR, and SCD (Fig. 7). Standardized regression coefficients revealed that mean hydrophobicity had a strong negative association with evolutionary variation in Ermin sequence properties, whose confidence intervals lay entirely below zero, indicating that reduced hydrophobicity independently characterizes Ermin evolution after accounting for correlated traits.

Fraction of charged residues (FCR) also showed a significant negative independent contribution, consistent with a modest reduction in overall charge density across vertebrate evolution. Sequence charge decoration (SCD) exhibited a significant positive independent effect, whereas sequence length and NCPR showed confidence intervals overlapping zero, indicating no independent contribution when considered jointly with other predictors.

Table 2. Root-to-tip PGLS trends in Ermin sequence evolution. Results of phylogenetic generalized least squares (PGLS) regressions assessing directional evolutionary trends in Ermin sequence properties across vertebrates. For each trait, the regression slope with respect to root-to-tip phylogenetic distance (Slope) and its associated p-value are reported, together with the maximum-likelihood estimate of Pagel's λ and the model log-likelihood (logLik). The log-likelihood denotes the maximum log-likelihood of the fitted PGLS model under the estimated phylogenetic covariance structure. Positive slope values indicate increasing trait values with increasing phylogenetic divergence from the root, whereas negative slope values indicate decreasing trait values. Because Kyte–Doolittle hydrophobicity values are expressed on a negative scale, positive slope values for hydrophobicity correspond to decreasing hydrophobicity over evolutionary divergence.

Trait	**N**	Pagel's λ	**logLik**	**Slope (root-to-tip)**	**p (slope)**
FCR	159	0.9999	428.8252	-0.0633	0.0152
NCPR	159	0.9999	409.018	0.0448	0.1293
SCD	159	0.9846	-508.504	17.4608	0.0598
Mean disorder	159	0.8732	287.877	-0.0825	0.1444
Mean hydrophobicity	159	0.9875	146.0139	0.504	0.0008

These results demonstrate that Ermin sequence evolution is primarily driven by compositional and patterning changes rather than by gross changes in protein length or net charge alone.

Taken together, the phylogeny-aligned trait mapping, clade-wise PGLS contrasts, root-to-tip analyses, and multivariate regression converge on a consistent evolutionary picture. Ermin sequences are intrinsically disordered across vertebrates, but disorder propensity is modulated in a lineage-specific, non-monotonic manner. Hydrophobicity shows a clearer directional decrease over evolutionary time, whereas electrostatic properties — especially charge patterning — undergo pronounced clade-specific restructuring. These results indicate that Ermin evolution is dominated by lineage-dependent tuning of sequence composition and electrostatic organization rather than by a simple, uniform increase in intrinsic disorder.

Discussion

Long IDRs as a conserved structural core of Ermin

The predominance of long intrinsically disordered regions across all vertebrate Ermin orthologs demonstrates that extensive disorder is a fundamental and conserved structural feature of the protein. Long IDRs are commonly associated with scaffolding functions, multivalent interactions, and regulatory flexibility (Wright and Dyson, 2015; Uversky, 2019), consistent with Ermin's role in cytoskeletal organization and oligodendrocyte biology (Zhang *et al.*, 2005; Meng *et al.*, 2010; Wang *et al.*, 2011). The near-universal presence of a C-terminal long IDR further suggests strong functional constraint, likely reflecting conserved interaction interfaces encompassing the characterized actin-binding segment (Zhang *et al.,* 2005; Ruskamo *et al.*, 2012) and potentially additional regulatory modules.

Fragmentation as an independent evolutionary dimension of disorder

By explicitly quantifying IDR fragmentation, we demonstrate that disorder organization evolves independently of overall disorder abundance. Birds represent a low-fragmentation regime in which long disordered regions remain largely continuous, whereas mammals exhibit a high-fragmentation architecture characterized by multiple long IDRs separated by ordered segments. This architectural distinction would not be detected by disorder fraction alone and highlights fragmentation as a distinct evolutionary parameter. These findings align with emerging views that IDP function depends critically on spatial organization and modular segmentation rather than solely on disorder content (van der Lee *et al.*, 2014; Darling and Uversky, 2018).

IDR amino acid composition: conserved constraints and lineage-specific diversification

Across vertebrates, IDR composition is shaped by conserved physicochemical constraints, reflected in consistent depletion of aromatic and bulky hydrophobic residues. The stability of these depletion patterns — particularly for tryptophan, phenylalanine, isoleucine, and leucine — supports the view that avoidance of structure-stabilizing chemistry represents a fundamental requirement for intrinsic disorder (Williams *et al.*, 2001; Lieutaud *et al.*, 2016; Oldfield *et al.*, 2019).

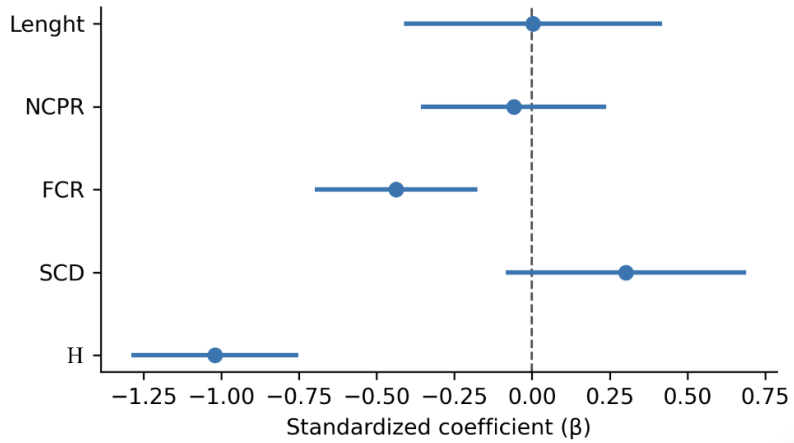


Figure 7. Multivariate PGLS regression coefficients for Ermin sequence properties. Standardized regression coefficients (β ± 95% confidence intervals) from a phylogenetic multiple regression including sequence length, fraction of charged residues (FCR), net charge per residue (NCPR), sequence charge decoration (SCD), and mean hydrophobicity (H) as predictors. The dashed vertical line indicates β = 0. Positive and negative coefficients represent independent contributions to evolutionary variation in Ermin sequence properties after accounting for phylogenetic relatedness. Pagel's λ = 0.93 corresponds to the maximum-likelihood estimate of phylogenetic signal for the fitted multivariate PGLS model.

In contrast, enrichment patterns vary markedly among clades, indicating that intrinsic disorder can be achieved through multiple compositional strategies. Early-diverging vertebrates rely predominantly on polar uncharged residues, whereas actinopterygians exhibit pronounced acidic enrichment. Amphibians and reptiles display increased compositional heterogeneity, while birds and mammals show the strongest low-complexity signatures. When analysed jointly, lineage-specific enrichments partially average out, revealing a conserved vertebrate IDR signature dominated by proline, glutamate, and polar residues, consistent with proteome-wide analyses of disordered regions (Djulbegovic and Uversky, 2022). Thus, vertebrate IDRs appear to be governed by universal biophysical constraints combined with clade-dependent compositional diversification.

Decoupling disorder from compositional evolution

A central outcome of this study is the decoupling of intrinsic disorder from compositional evolutionary trends. Root-to-tip analyses reveal a strong directional decrease in mean hydrophobicity and a weaker but statistically significant reduction in overall charge density (FCR) across vertebrate evolution, whereas disorder propensity and sequence charge decoration (SCD) do not

exhibit consistent monotonic trends. Multivariate analyses further demonstrate that hydrophobicity, global charge density, and charge patterning independently contribute to evolutionary divergence.

Because Kyte–Doolittle hydrophobicity values for disordered proteins are typically negative, shifts toward more negative values reflect progressive selection against hydrophobic residues rather than increased structural order. The concurrent reduction in global charge density suggests gradual compositional retuning, while the independent effect of SCD indicates lineage-specific restructuring of charge distribution along the sequence. Together, these results show that Ermin maintains a conserved disordered scaffold while undergoing directional compositional shifts and clade-specific electrostatic remodelling.

Phylogenetic structuring of sequence evolution

High Pagel's λ values across traits indicate substantial phylogenetic signal, confirming that variation in Ermin sequence properties is strongly structured by shared ancestry. Importantly, strong phylogenetic signal does not preclude adaptation; rather, it indicates that evolutionary changes accumulate in a lineage-dependent manner such that closely related species resemble one another more than expected under independence (Pagel, 1999; Blomberg et al., 2003). Although clade-wise distributions and phylogeny-aligned heatmaps visually suggest lineage-specific differences, formal PGLS likelihood-ratio tests indicate that inclusion of categorical clade identity does not significantly improve model fit once phylogenetic covariance is accounted for. Similarly, pairwise PGLS contrasts do not remain significant after conservative multiple-testing correction. These findings suggest that variation in Ermin sequence architecture is not partitioned into statistically discrete clade-level regimes but instead reflects gradual, phylogenetically structured divergence along evolutionary lineages. Thus, the observed clustering in heatmaps and trait distributions is best interpreted as the outcome of shared evolutionary history and distributed lineage-dependent tuning rather than abrupt evolutionary shifts confined to particular vertebrate clades. Ermin sequence evolution therefore appears to proceed through incremental modifications along the phylogeny rather than through categorical regime changes.

Functional context: Ermin among myelin-associated IDPs

Placing Ermin within the broader context of myelin-associated intrinsically disordered proteins underscores its functional distinctiveness. Classical myelin proteins such as myelin basic protein (MBP) and the P0 cytoplasmic tail undergo partial folding upon membrane association and act as molecular

adhesives within compact myelin (Boggs, 2006; Harauz and Musse, 2007). In contrast, Ermin functions primarily as an actin-associated cytoskeletal regulator, favoring sustained disorder, low hydrophobicity, and finely tuned electrostatic organization. The evolutionary patterns observed here — directional reduction in hydrophobicity, modest decrease in global charge density, and lineage-specific modulation of charge patterning — are consistent with a highly dynamic interaction scaffold rather than a membrane-anchored structural protein.

Evolutionary implications

Collectively, phylogenetic mapping, root-to-tip regressions, and multivariate PGLS analyses converge on a model in which Ermin evolution is dominated by gradual, lineage-structured compositional tuning of an already intrinsically disordered scaffold. Intrinsic disorder itself is deeply conserved across vertebrates and does not exhibit a uniform directional increase. Instead, early-diverging lineages already display high disorder levels, indicating that extensive intrinsic disorder represents an ancestral structural state rather than a derived innovation.

In contrast, mean hydrophobicity shows a consistent directional shift along the phylogeny, reflecting progressive selection against hydrophobic residues over vertebrate divergence. Electrostatic properties, particularly charge patterning, vary in a phylogenetically structured but non-discrete manner, suggesting distributed modulation of ensemble behavior rather than categorical regime shifts among major clades. The absence of statistically separable clade effects indicates that evolutionary change in Ermin is best understood as incremental divergence accumulating along branches rather than as abrupt transitions at major taxonomic boundaries.

These findings reinforce the view that intrinsic disorder is an emergent property shaped by multiple interacting sequence features, including hydrophobicity, charge density, and charge patterning, rather than by a single evolutionary axis. Ermin evolution therefore exemplifies how intrinsically disordered proteins diversify through subtle compositional redistribution and electrostatic retuning while maintaining a conserved disordered structural framework. By integrating phylogenetic comparative methods with polymer- and disorder-based metrics, this study provides a quantitative framework for understanding how IDPs evolve across deep evolutionary timescales through gradual lineage-dependent modification rather than discrete structural innovation.

Conclusion

This study provides a comprehensive phylogenetic and sequence-level analysis of 159 vertebrate Ermin orthologs and establishes Ermin as a deeply conserved intrinsically disordered protein. Across all examined clades, Ermin consistently meets criteria for high intrinsic disorder, indicating that extensive disorder represents an ancestral and fundamental structural feature rather than a lineage-specific innovation.

Despite this conserved disordered framework, Ermin evolution is characterized by lineage-dependent remodeling of disorder architecture and composition. Mammals exhibit increased fragmentation of long intrinsically disordered regions (IDRs), whereas birds maintain largely continuous long-IDR architectures, demonstrating that IDR partitioning evolves independently of overall disorder abundance. Amino-acid composition analyses reveal conserved depletion of aromatic and bulky hydrophobic residues, combined with clade-structured variation in charged and disorder-promoting residues.

Phylogenetic comparative analyses further indicate a directional decrease in mean hydrophobicity across vertebrate divergence, consistent with progressive selection against hydrophobic residues within an already disordered scaffold. In contrast, disorder propensity and charge patterning do not exhibit uniform monotonic trends, and formal clade-wise tests do not support discrete categorical regime shifts after accounting for phylogenetic covariance. Instead, evolutionary variation in Ermin sequence properties is best explained by gradual, lineage-structured divergence along the phylogeny.

Together, these findings support a model in which Ermin evolution proceeds through incremental architectural restructuring and electrostatic retuning of a conserved intrinsically disordered framework. This highlights IDR fragmentation, compositional redistribution, and charge organization as key evolutionary dimensions shaping diversification of intrinsically disordered proteins across vertebrate lineages.

References

Aggarwal, S., Yurlova, L. & Simons, M., 2011. Central nervous system myelin: structure, synthesis and assembly. *Trends Cell Biol.* 21, 585–593. https://doi.org/10.1016/j.tcb.2011.06.004

Bah, A. & Forman-Kay, J.D. 2016. Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. *J. Biol. Chem.* 291, 6696–6705. https://doi.org/10.1074/jbc.R115.695056

Baumann, N. & Pham-Dinh, D., 2001. Biology of Oligodendrocyte and Myelin in the Mammalian Central Nervous System. *Physiol. Rev.* 81, 871–927. https://doi.org/10.1152/physrev.2001.81.2.871

Bellay, J., Han, S., Michaut, M., Kim, T., Costanzo, M., Andrews, B.J., Boone, C., Bader, G.D., Myers, C.L. & Kim, P.M., 2011. Bringing order to protein disorder through comparative genomics and genetic interactions. *Genome Biol.* 12, R14. https://doi.org/10.1186/gb-2011-12-2-r14

Blomberg, S.P., Garland, T. & Ives, A.R., 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57, 717–745. https://doi.org/10.1111/j.0014-3820.2003.tb00285.x

Boggs, J.M., 2006. Myelin basic protein: a multifunctional protein. *Cell. Mol. Life Sci.* 63, 1945–1961. https://doi.org/10.1007/s00018-006-6094-7

Brockschnieder, D., Sabanay, H., Riethmacher, D. & Peles, E., 2006. Ermin, A Myelinating Oligodendrocyte-Specific Protein That Regulates Cell Morphology. *J. Neurosci.* 26, 757–762. https://doi.org/10.1523/JNEUROSCI.4317-05.2006

Burnham, K.P. & Anderson, D.R., 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd ed. Springer, New York. http://doi.org/10.1007/b97636

Cooper, N., Thomas, G.H., Venditti, C., Meade, A. & Freckleton, R.P., 2016. A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. *Biol. J. Linn. Soc.* 118, 64–77. https://doi.org/10.1111/bij.12701

Darling, A.L. & Uversky, V.N., 2018. Intrinsic Disorder and Posttranslational Modifications: The Darker Side of the Biological Dark Matter. *Front. Genet.* 9, 158. https://doi.org/10.3389/fgene.2018.00158

Das, R.K. & Pappu, R.V., 2013. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U.S.A.* 110, 13392–13397. https://doi.org/10.1073/pnas.1304749110

Das, R.K., Ruff, K.M. & Pappu, R.V., 2015. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 32, 102–112. https://doi.org/10.1016/j.sbi.2015.03.008

Dayhoff, G.W. & Uversky, V.N., 2022. Rapid prediction and analysis of protein intrinsic disorder. *Protein Sci.* 31, e4496. https://doi.org/10.1002/pro.4496

Djulbegovic, M. & Uversky, V.N., 2022. The aqueous humor proteome is intrinsically disordered. Biochemistry and *Biophys. Rep.* 29, 101202. https://doi.org/10.1016/j.bbrep.2022.101202

Dosztányi, Z., Csizmók, V., Tompa, P. & Simon, I., 2005. The Pairwise Energy Content Estimated from Amino Acid Composition Discriminates between Folded and Intrinsically Unstructured Proteins. *J. Mol. Biol.* 347, 827–839. https://doi.org/10.1016/j.jmb.2005.01.071

Erdős, G. & Dosztányi, Z., 2020. Analyzing Protein Disorder with IUPred2A. *Curr. Prot. Bioinform.* 70, e99. https://doi.org/10.1002/cpbi.99

Farris, J.S., 1972. Estimating Phylogenetic Trees from Distance Matrices. *Am. Nat.* 106, 645–668. https://doi.org/10.1086/282802

Freckleton, R.P., Harvey, P.H. & Pagel, M., 2002. Phylogenetic Analysis and Comparative Data: A Test and Review of Evidence. *Am. Nat.* 160, 712–726. https://doi.org/10.1086/343873

Ghosh, K., Huihui, J., Phillips, M. & Haider, A., 2022. Rules of Physical Mathematics Govern Intrinsically Disordered Proteins. *Annu. Rev. Biophys.* 51, 355–376. https://doi.org/10.1146/annurev-biophys-120221-095357

Grafen, A., 1989. The phylogenetic regression. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 326, 119–157. https://doi.org/10.1098/rstb.1989.0106

Gsponer, J., Futschik, M.E., Teichmann, S.A. & Babu, M.M., 2008. Tight Regulation of Unstructured Proteins: From Transcript Synthesis to Protein Degradation. *Science 322, 1365–1368*. https://doi.org/10.1126/science.1163581

Harauz, G. & Boggs, J.M., 2013. Myelin management by the 18.5-kDa and 21.5-kDa classic myelin basic protein isoforms. *J. Neurochem.* 125, 334–361. https://doi.org/10.1111/jnc.12195

Harauz, G. & Musse, A.A., 2007. A Tale of Two Citrullines—Structural and Functional Aspects of Myelin Basic Protein Deimination in Health and Disease. *Neurochem. Res.* 32, 137–158. https://doi.org/10.1007/s11064-006-9108-9

Hess, P.N. & De Moraes Russo, C.A., 2007. An empirical test of the midpoint rooting method: a test for the midpoint rooting method. *Biol. J. Linn. Soc.* 92, 669–674. https://doi.org/10.1111/j.1095-8312.2007.00864.x

Holm, S., 1979. A Simple Sequentially Rejective Multiple Test Procedure. *Scand. J. Stat.* 6, 65–70. https://doi.org/10.2307/4615733

Holehouse, A.S., & Kragelund, B.B., 2024. The molecular basis for cellular function of intrinsically disordered protein regions. *Nat. Rev. Mol. Cell. Biol.* 25, 187–211. https://doi.org/10.1038/s41580-023-00673-0

Homs, A., Codina-Solà, M., Rodríguez-Santiago, B., Villanueva, C.M., Monk, D., Cuscó, I. & Pérez-Jurado, L.A., 2016. Genetic and epigenetic methylation defects and implication of the ERMN gene in autism spectrum disorders. *Transl. Psychiatry* 6, e855–e855. https://doi.org/10.1038/tp.2016.120

Hunt, G., 2006. Fitting and comparing models of phyletic evolution: random walks and beyond. *Paleobiology* 32, 578–601. https://doi.org/10.1666/05070.1

Jahn, O., Tenzer, S. & Werner, H.B., 2009. Myelin Proteomics: Molecular Anatomy of an Insulating Sheath. *Mol. Neurobiol.* 40, 55–72. https://doi.org/10.1007/s12035-009-8071-2

Koszorus, I. & Kósa, F., 2025. Evolutionary patterns of structural disorder and post-translational modifications in the 18.5 kDa myelin basic protein. Studia UBB Biologia 70, 115–150. https://doi.org/10.24193/subbbiol.2025.2.08

Kragelund, B.B., Jensen, M.K. & Skriver, K., 2012. Order by disorder in plant signaling. *Trends. Plant. Sci.* 17, 625–632. https://doi.org/ 10.1016/j.tplants.2012.06.010

Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K., 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* 35, 1547–1549. https://doi.org/10.1093/molbev/msy096

Kyte, J. & Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105–132. https://doi.org/10.1016/0022-2836(82)90515-0

Liang F, Hwang JH, Tang NW & Hunziker W., 2018. Juxtanodin in retinal pigment epithelial cells: Expression and biological activities in regulating cell morphology and actin cytoskeleton organization. *J. Comp. Neurol.* 526(2):205-215. https://doi.org/10.1002/cne.24301

Li X, Romero P, Rani M, Dunker AK & Obradovic Z. Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome Inform. Ser. Workshop Genome Inform.* 1999;10:30-40. PMID: 11072340.

Lieutaud, P., Ferron, F., Uversky, A.V., Kurgan, L., Uversky, V.N. & Longhi, S., 2016. How disordered is my protein and what is its disorder for? A guide through the "dark side" of the protein universe. *Intrinsically Disord. Proteins* 4, e1259708. https://doi.org/10.1080/21690707.2016.1259708

Martins, E.P. & Hansen, T.F., 1997. Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am. Nat.* 149, 646–667. https://doi.org/10.1086/286013

Meng J, Xia W, Tang J, Tang BL & Liang F., 2010. Dephosphorylation-dependent inhibitory activity of juxtanodin on filamentous actin disassembly. *J. Biol. Chem.* 10;285(37):28838-49 https://doi.org/10.1074/jbc.M110.117887

Monzon, A.M., Necci, M., Quaglia, F., Walsh, I., Zanotti, G., Piovesan, D. & Tosatto, S.C.E., 2020. Experimentally Determined Long Intrinsically Disordered Protein Regions Are Now Abundant in the Protein Data Bank. *IJMS.* 21, 4496. https://doi.org/10.3390/ijms21124496

Nave, K.-A. & Werner, H.B., 2014. Myelination of the Nervous System: Mechanisms and Functions. *Annu. Rev. Cell Dev. Biol.* 30, 503–533. https://doi.org/10.1146/annurev-cellbio-100913-013101

Nawaz, S., Sánchez, P., Schmitt, S., Snaidero, N., Mitkovski, M., Velte, C., Brückner, B.R., Alexopoulos, I., Czopka, T., Jung, S.Y., Rhee, J.S., Janshoff, A., Witke, W., Schaap, I.A.T., Lyons, D.A. & Simons, M., 2015. Actin Filament Turnover Drives Leading Edge Growth during Myelin Sheath Formation in the Central Nervous System. *Dev. Cell* 34, 139–151. https://doi.org/10.1016/j.devcel.2015.05.013

Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C.J. & Dunker, A.K., 2003. Predicting intrinsic disorder from amino acid sequence. *Proteins* 53, 566–572. https://doi.org/10.1002/prot.10532

Oldfield, C.J., Uversky, V.N., Dunker, A.K. & Kurgan, L., 2019. Introduction to intrinsically disordered proteins and regions, in: Intrinsically Disordered Proteins. *Elsevier*, pp. 1–34. https://doi.org/10.1016/B978-0-12-816348-1.00001-6

Pagel, M., 1999. Inferring the historical patterns of biological evolution. *Nature* 401, 877–884. https://doi.org/10.1038/44766

Peng, K., Radivojac, P., Vucetic, S., Dunker, A.K. & Obradovic, Z., 2006. Length-dependent prediction of protein intrinsic disorder. *BMC. Bioinformatics* 7, 208. https://doi.org/10.1186/1471-2105-7-208

Peng, Z., Yan, J., Fan, X., Mizianty, M.J., Xue, B., Wang, K., Hu, G., Uversky, V.N. & Kurgan, L., 2015. Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell. Mol. Life Sci.* 72, 137–151. https://doi.org/10.1007/s00018-014-1661-9

Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H. & Ferrin, T.E., 2021. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. 30, 70–82. https://doi.org/10.1002/pro.3943

Raasakka, A. & Kursula, P., 2020. Flexible Players within the Sheaths: The Intrinsically Disordered Proteins of Myelin in Health and Disease. *Cells* 9, 470. https://doi.org/10.3390/cells9020470

Radivojac, P., Obradovic, Z., Brown, C.J. & Dunker A.K., 2003. Prediction of boundaries between intrinsically ordered and disordered protein regions. *Pac. Symp. Biocomput*. 216–227

Rajagopalan, K., Mooney, S.M., Parekh, N., Getzenberg, R.H. & Kulkarni, P., 2011. A majority of the cancer/testis antigens are intrinsically disordered proteins. *J. Cell. Biochem. 112, 3256–3267.* https://doi.org/10.1002/jcb.23252

Ruskamo, S., Chukhlieb, M., Vahokoski, J., Bhargav, S.P., Liang, F., Kursula, I. & Kursula, P., 2012. Juxtanodin is an intrinsically disordered F-actin-binding protein. *Sci. Rep.* 2, 899. https://doi.org/10.1038/srep00899

Salek Esfahani, B., Gharesouran, J., Ghafouri-Fard, S., Talebian, S., Arsang-Jang, S., Omrani, M.D., Taheri, M. & Rezazadeh, M., 2019. Down-regulation of ERMN expression in relapsing remitting multiple sclerosis. *Metab. Brain Dis.* 34, 1261–1266. https://doi.org/10.1007/s11011-019-00429-w

Sawle, L. & Ghosh, K., 2015. A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins. *J. Chem. Phys.* 143, 085101. https://doi.org/10.1063/1.4929391

Shiva, S., Gharesouran, J., Sabaie, H., Asadi, M.R., Arsang-Jang, S., Taheri, M. & Rezazadeh, M., 2021. Expression Analysis of Ermin and Listerin E3 Ubiquitin Protein Ligase 1 Genes in Autistic Patients. *Front. Mol. Neurosci.* 14, 701977. https://doi.org/10.3389/fnmol.2021.701977

Saitou N. & Nei M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406-425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Uversky, V.N., 2019. Intrinsically Disordered Proteins and Their “Mysterious” (Meta)Physics. *Front. Phys.* 7, 10. https://doi.org/10.3389/fphy.2019.00010

Uversky, V.N., Gillespie, J.R. & Fink, A.L., 2000. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins* 41, 415–427. https://doi.org/10.1002/1097-0134(20001115)41:3<415::AID-PROT130>3.0.CO;2-7

Vacic, V. & Iakoucheva, L.M., 2012. Disease mutations in disordered regions—exception to the rule? *Mol. BioSyst.* 8, 27–32. https://doi.org/10.1039/C1MB05251A

Vacic, V., Markwick, P.R.L., Oldfield, C.J., Zhao, X., Haynes, C., Uversky, V.N. & Iakoucheva, L.M., 2012. Disease-Associated Mutations Disrupt Functionally Important Regions of Intrinsic Protein Disorder. *PLoS. Comput. Biol.* 8, e1002709. https://doi.org/10.1371/journal.pcbi.1002709

Vacic, V., Uversky, V.N., Dunker, A.K. & Lonardi, S., 2007. Composition Profiler: a tool for discovery and visualization of amino acid composition differences. *BMC. Bioinformatics* 8, 211. https://doi.org/10.1186/1471-2105-8-211

Van Der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., Kim, P.M., Kriwacki, R.W., Oldfield, C.J., Pappu, R.V., Tompa, P., Uversky, V.N., Wright, P.E. & Babu, M.M., 2014. Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* 114, 6589–6631. https://doi.org/10.1021/cr400525m

Wang, S., Wang, T., Liu, T., Xie, R., Zhao, X., Wang, L., Yang, Q., Jia, L. & Han, J., 2020. Ermin is a p116^{RIP} -interacting protein promoting oligodendroglial differentiation and myelin maintenance. *Glia* 68, 2264–2276. https://doi.org/10.1002/glia.23838

Wang, T., Jia, L., Lv, B., Liu, B., Wang, W., Wang, F., Yang, G., Bu, X., Yao, L. & Zhang, B., 2011. Human Ermin (hErmin), a new oligodendrocyte-specific cytoskeletal protein related to epileptic seizure. *Brain Res.* 1367, 77–84. https://doi.org/10.1016/j.brainres.2010.10.003

Williams, R.M., Obradovic, Z., Mathura, V., Braun, W., Garner, E.C., Young, J., Takayama, S., Brown, C.J. & Dunker, A.K., 2001. The protein non-folding problem: Amino acid determinants of intrinsic order and disorder. In: *Biocomputing 2001*. Proceedings of the Pacific Symposium on Biocomputing. World Scientific, Mauna Lani, Hawaii, pp. 89–100. https://doi.org/10.1142/9789814447362_0010

Wright, P.E. & Dyson, H.J., 2015. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* 16, 18–29. https://doi.org/10.1038/nrm3920

Wright, P.E. & Dyson, H.J., 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 293, 321–331. https://doi.org/10.1006/jmbi.1999.3110

Xue, B., Dunbrack, R.L., Williams, R.W., Dunker, A.K. & Uversky, V.N., 2010. PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta Proteins Proteom.* 1804, 996–1010. https://doi.org/10.1016/j.bbapap.2010.01.011

Zhang, B., Cao, Q., Guo, A., Chu, H., Chan, Y.G., Buschdorf, J.P., Low, B.C., Ling, E.A. & Liang, F., 2005. Juxtanodin: An oligodendroglial protein that promotes cellular arborization and 2′,3′-cyclic nucleotide-3′-phosphodiesterase trafficking. *Proc. Natl. Acad. Sci. U.S.A.* 102, 11527–11532. https://doi.org/10.1073/pnas.0500952102

Ziaei, A., Garcia-Miralles, M., Radulescu, C.I., Sidik, H., Silvin, A., Bae, H., Bonnard, C., Yusof, N.A.B.M., Ferrari Bardile, C., Tan, L.J., Ng, A.Y.J., Tohari, S., Dehghani, L., Henry, L., Yeo, X.Y., Lee, S., Venkatesh, B., Langley, S.R., Shaygannejad, V., Reversade, B., Jung, S., Ginhoux, F. & Pouladi, M.A., 2022. Ermin deficiency leads to compromised myelin, inflammatory milieu, and susceptibility to demyelinating insult. *Brain Path.* 32(5):e13064. https://doi.org/10.1111/bpa.13064

Supplementary materials

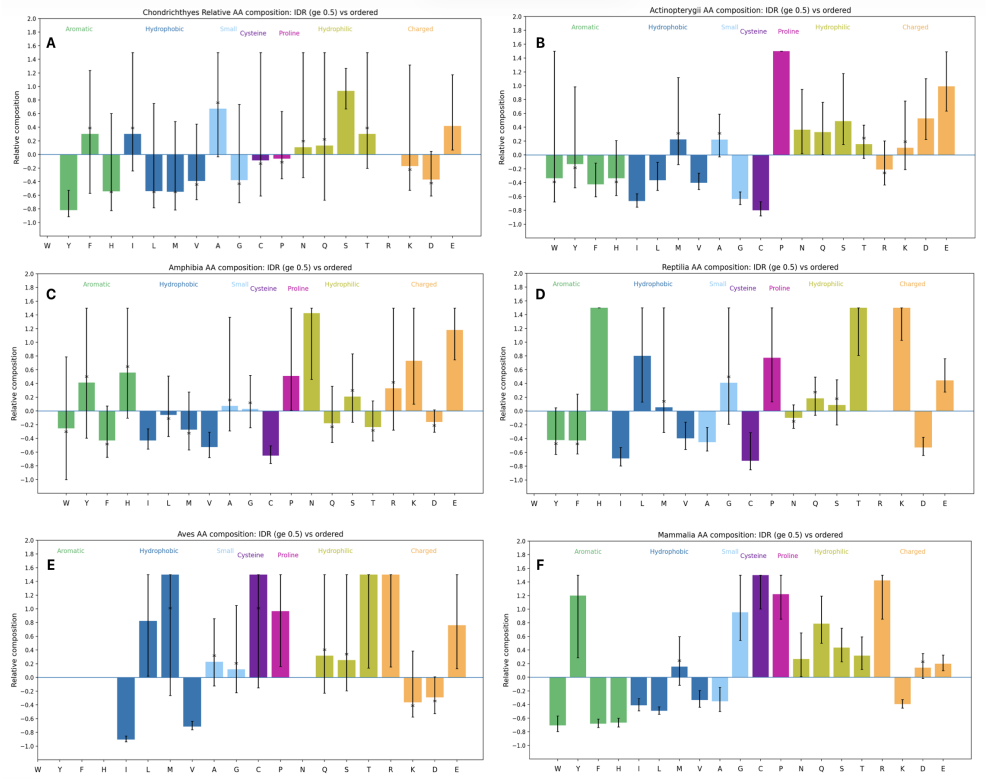


Figure S1. Relative amino-acid composition of intrinsically disordered regions (IDRs) versus ordered regions in Ermin proteins. A. *Chondrichthyes*, B. *Actinopterygii*, C. *Amphibia*, D. *Reptilia*, E. *Aves*, F. *Mammalia.* Relative composition was calculated as $(f_{IDR} - f_{ordered}) / f_{ordered}$; positive values indicate enrichment in IDRs and negative values indicate depletion. Amino acids are grouped by physicochemical properties, with cysteine and proline shown separately. Error bars represent 95% bootstrap confidence intervals (10,000 sequence resamplings). Bootstrap resampling was performed at the sequence level. Asterisks denote residues whose confidence intervals overlap zero (non-significant enrichment or depletion).

Table S1. Pairwise PGLS clade contrasts for Ermin sequence properties.
Differences represent estimated contrasts between clade means while accounting for phylogenetic covariance. Column definitions: trait — trait analyzed (FCR, NCPR, SCD, disorder_mean, mean_hydrophobicity_KD); Clade A — first clade in the comparison; Clade B — second clade in the comparison; Mean difference (A − B) — estimated difference between clade means from the fitted PGLS model (effect size in the original trait units); p (raw) — unadjusted p-value for the clade contrast; p (Holm-adjusted) — Holm–Bonferroni–adjusted p-value controlling the family-wise error rate across pairwise contrasts. Note: Positive values of Mean difference (A − B) indicate higher trait values in Clade A relative to Clade B; negative values indicate the opposite. No pairwise comparisons remained significant after Holm-Bonferroni correction.

Trait	Clade A	Clade B	Mean difference (A − B)	p (raw)	p (Holm-adjusted)
FCR	*Actinopterygii*	*Chondrichthyes*	0.0601	0.0414	0.6215
FCR	*Chondrichthyes*	*Mammalia*	-0.0735	0.0424	0.6215
FCR	*Aves*	*Mammalia*	-0.0456	0.1152	1.0000
FCR	*Amphibia*	*Mammalia*	-0.0423	0.1750	1.0000
FCR	*Chondrichthyes*	*Reptilia*	-0.0446	0.1873	1.0000
FCR	*Actinopterygii*	*Amphibia*	0.0289	0.2748	1.0000
FCR	*Actinopterygii*	*Aves*	0.0322	0.2903	1.0000
FCR	*Amphibia*	*Chondrichthyes*	0.0312	0.3168	1.0000
FCR	*Mammalia*	*Reptilia*	0.0290	0.3446	1.0000
FCR	*Aves*	*Chondrichthyes*	0.0279	0.4197	1.0000
FCR	*Aves*	*Reptilia*	-0.0166	0.5560	1.0000
FCR	*Actinopterygii*	*Reptilia*	0.0156	0.5994	1.0000
FCR	*Amphibia*	*Reptilia*	-0.0134	0.6329	1.0000
FCR	*Actinopterygii*	*Mammalia*	-0.0134	0.6830	1.0000
FCR	*Amphibia*	*Aves*	0.0033	0.9093	1.0000
NCPR	*Amphibia*	*Mammalia*	-0.0474	0.2044	1.0000
NCPR	*Actinopterygii*	*Mammalia*	-0.0468	0.2333	1.0000
NCPR	*Aves*	*Mammalia*	-0.0375	0.2787	1.0000
NCPR	*Amphibia*	*Reptilia*	-0.0340	0.3093	1.0000
NCPR	*Actinopterygii*	*Reptilia*	-0.0334	0.3458	1.0000
NCPR	*Aves*	*Reptilia*	-0.0241	0.4751	1.0000
NCPR	*Chondrichthyes*	*Mammalia*	-0.0289	0.5046	1.0000
NCPR	*Actinopterygii*	*Chondrichthyes*	-0.0179	0.6124	1.0000
NCPR	*Amphibia*	*Chondrichthyes*	-0.0185	0.6205	1.0000
NCPR	*Chondrichthyes*	*Reptilia*	-0.0155	0.7004	1.0000
NCPR	*Mammalia*	*Reptilia*	0.0134	0.7155	1.0000
NCPR	*Amphibia*	*Aves*	-0.0099	0.7752	1.0000
NCPR	*Actinopterygii*	*Aves*	-0.0093	0.7993	1.0000
NCPR	*Aves*	*Chondrichthyes*	-0.0086	0.8353	1.0000
NCPR	*Actinopterygii*	*Amphibia*	0.0006	0.9851	1.0000
SCD	*Actinopterygii*	*Mammalia*	16.1180	0.1260	1.0000
SCD	*Amphibia*	*Mammalia*	14.1134	0.1590	1.0000

Trait	Clade A	Clade B	Mean difference (A − B)	p (raw)	p (Holm-adjusted)
SCD	*Actinopterygii*	*Reptilia*	12.9310	0.1742	1.0000
SCD	*Amphibia*	*Reptilia*	10.9264	0.2237	1.0000
SCD	*Actinopterygii*	*Aves*	8.9645	0.3594	1.0000
SCD	*Chondrichthyes*	*Mammalia*	10.0325	0.3886	1.0000
SCD	*Aves*	*Mammalia*	7.1535	0.4416	1.0000
SCD	*Amphibia*	*Aves*	6.9599	0.4526	1.0000
SCD	*Actinopterygii*	*Chondrichthyes*	6.0855	0.5214	1.0000
SCD	*Chondrichthyes*	*Reptilia*	6.8455	0.5285	1.0000
SCD	*Aves*	*Reptilia*	3.9665	0.6621	1.0000
SCD	*Amphibia*	*Chondrichthyes*	4.0809	0.6839	1.0000
SCD	*Mammalia*	*Reptilia*	-3.1870	0.7460	1.0000
SCD	*Aves*	*Chondrichthyes*	-2.8790	0.7957	1.0000
SCD	*Actinopterygii*	*Amphibia*	2.0046	0.8137	1.0000
mean disorder	*Aves*	*Mammalia*	0.1183	0.0222	0.3326
mean disorder	*Amphibia*	*Aves*	-0.1119	0.0306	0.4282
mean disorder	*Aves*	*Reptilia*	0.0880	0.0819	1.0000
mean disorder	*Actinopterygii*	*Aves*	-0.0651	0.2316	1.0000
mean disorder	*Aves*	*Chondrichthyes*	0.0656	0.2957	1.0000
mean disorder	*Actinopterygii*	*Amphibia*	0.0468	0.3254	1.0000
mean disorder	*Actinopterygii*	*Mammalia*	0.0532	0.3629	1.0000
mean disorder	*Amphibia*	*Chondrichthyes*	-0.0462	0.4179	1.0000
mean disorder	*Chondrichthyes*	*Mammalia*	0.0527	0.4190	1.0000
mean disorder	*Mammalia*	*Reptilia*	-0.0303	0.5793	1.0000
mean disorder	*Amphibia*	*Reptilia*	-0.0239	0.6342	1.0000
mean disorder	*Actinopterygii*	*Reptilia*	0.0229	0.6650	1.0000
mean disorder	*Chondrichthyes*	*Reptilia*	0.0224	0.7154	1.0000
mean disorder	*Amphibia*	*Mammalia*	0.0064	0.9083	1.0000
mean disorder	*Actinopterygii*	*Chondrichthyes*	0.0006	0.9917	1.0000
mean hydrophobicity	*Chondrichthyes*	*Reptilia*	0.4230	0.0135	0.2028
mean hydrophobicity	*Chondrichthyes*	*Mammalia*	0.4140	0.0238	0.3328
mean hydrophobicity	*Actinopterygii*	*Chondrichthyes*	-0.3015	0.0448	0.5830
mean hydrophobicity	*Amphibia*	*Reptilia*	0.2816	0.0461	0.5830
mean hydrophobicity	*Amphibia*	*Mammalia*	0.2726	0.0834	0.9176
mean hydrophobicity	*Aves*	*Reptilia*	0.2403	0.0921	0.9214
mean hydrophobicity	*Aves*	*Mammalia*	0.2313	0.1134	1.0000
mean hydrophobicity	*Actinopterygii*	*Amphibia*	-0.1600	0.2317	1.0000
mean hydrophobicity	*Aves*	*Chondrichthyes*	-0.1827	0.2972	1.0000
mean hydrophobicity	*Amphibia*	*Chondrichthyes*	-0.1415	0.3717	1.0000
mean hydrophobicity	*Actinopterygii*	*Reptilia*	0.1216	0.4162	1.0000
mean hydrophobicity	*Actinopterygii*	*Aves*	-0.1187	0.4397	1.0000
mean hydrophobicity	*Actinopterygii*	*Mammalia*	0.1126	0.4964	1.0000
mean hydrophobicity	*Amphibia*	*Aves*	0.0413	0.7768	1.0000
mean hydrophobicity	*Mammalia*	*Reptilia*	0.0090	0.9536	1.0000

Table S2. Likelihood-ratio tests of clade effects in univariate PGLS models. Summary of univariate PGLS models testing whether vertebrate clade identity improves model fit for each Ermin sequence trait. For each trait, a full model including clade identity was compared to a reduced (intercept-only) model using a likelihood-ratio test (LRT). Pagel's λ was estimated by maximum likelihood for both full and reduced models. Row definitions: Trait — analyzed sequence property (mean disorder, mean hydrophobicity, FCR, NCPR, or SCD); n — number of species included in the model; Pagel's λ (full) — maximum-likelihood estimate of phylogenetic signal in the full model; logLik (full) — log-likelihood of the model including clade identity; logLik (reduced) — log-likelihood of the intercept-only model; LR statistic — likelihood-ratio statistic calculated as 2 × (logLik_full − logLik_reduced); df — degrees of freedom corresponding to the number of clade parameters; p (LRT, clade effect) — p-value of the likelihood-ratio test evaluating the significance of clade identity. Significant LRT p-values indicate that clade membership explains trait variation beyond phylogenetic covariance alone.

Trait	FCR	NCPR	SCD	Mean disorder	Mean hydrophobicity
n	159	159	159	159	159
λ (full model)	1.0000	1.0000	0.9919	0.8516	0.9633
logLik (full model)	437.6059	409.2020	-484.9054	290.5840	162.0218
logLik (reduced model)	434.0968	407.8885	-486.6154	287.1705	156.9697
LR statistic	7.0182	2.6270	3.4201	6.8270	10.1044
df	5	5	5	5	5
p (LRT, clade effect)	0.2193	0.7573	0.6355	0.2338	0.0723

STUDIA UNIVERSITATIS BABEŞ-BOLYAI BIOLOGIA, 71, 1, 2026 (p. 161-184)
doi:10.24193/subbbiol.2026.1.08

Evaluating soil microbial metabolic activity in response to two plant biostimulants using the Biolog Ecoplate™ method

Katya Dimitrova

Agricultural University – Plovdiv, Plovdiv, Bulgaria
E-mail: katia_dimitrova@au-plovdiv.bg

Article history: Received 16 February 2026; Revised 17 June 2026; Accepted 19 June 2026; Available online 25 June 2026

Abstract. Biostimulants are emerging as viable alternatives to chemical fertilizers, offering a pathway toward more sustainable and environmentally friendly agricultural practices. Their effectiveness, however, is influenced by a range of environmental and biological conditions, making outcomes variable. Evidence suggests that biostimulants may exert their influence by modifying the metabolic activity and composition of soil microbial communities – key drivers of nutrient cycling. This study aimed to assess microbial activity in untreated soil and soil treated with two distinct biostimulants: a seaweed-derived product (PStim) and a microbial-based formulation (MbPB), under greenhouse conditions. Microbial dynamics were evaluated using the Biolog EcoPlate™ method, focusing on substrate utilization pattern. Soil respiration rate was also estimated. No significant changes in soil chemical properties were observed following biostimulant application. Nonetheless, all variants demonstrated high metabolic activity, as indicated by average well-color development (AWCD), with notable differences in substrate preferences. Functional indices pointed to increased metabolic diversity and evenness in biostimulant-treated soils, potentially linked to the targeted stimulation of specific microbial taxa.

Keywords: soil microbial activity, Biolog EcoPlate™, plant biostimulants, microbial biostimulants, substrate utilization patterns, functional indices

Introduction

Agriculture is of paramount importance for providing food, materials, and resources for industry and society. However, due to its immense intensification and the adoption of environmentally detrimental practices, it is currently considered responsible for a third of greenhouse gas emissions, 80% of deforestation, and 70% of terrestrial biodiversity loss (FAO, 2021). It is expected that the problems along the food chain and its social impact will increase in the future and warrant efforts and action in order to address sustainability challenges (Pe'er *et al*., 2020; Verfuerth *et al*., 2023). Among the European Union tasks to alleviate the negative effects of agricultural practices is the reduction of mineral fertilizers with 20% by 2050. Such policies aim to stimulate the transition from contemporary, highly intensified agriculture towards adoption of more eco-friendly approaches and provision of sustainability. The concept of sustainable agriculture is based on four main principles: land management, resource management, human interaction, and the ecosystem interface (Shelef *et al*., 2018). Sustainable agriculture is a holistic approach which purpose is to meet the current food production needs while preserving environment, support economic viability, and ensure the social equity for future generations. In general, it involves practices that promote soil health, biodiversity, water conservation, and energy efficiency, while minimizing the use of synthetic inputs. The assessment of various approaches toward sustainable agriculture and their potential short- or long-term impact on the environment, economy, society, and politics is underway (Zhang and Drury, 2024). One of the possible approaches of reducing reliance on chemical fertilizers is the application of plant biostimulants.

Plant biostimulants encompass substances and microorganisms aimed at enhancing natural processes within plants, improving nutrient uptake and efficiency, stress tolerance, and overall crop quality (EU Regulation 2019/100). The main substances in the biostimulants are humic and fulvic acids, seaweed extracts, protein hydrolysates, chitosan and other biopolymers, inorganic compounds and beneficial microorganism (Tejada *et al*., 2011; Pichyangkura and Chadchawan, 2015; Tiwari *et al*., 2019; Rathor *et al*., 2023; Mughunth *et al*., 2024). The application of biostimulants in horticulture is increasing due to their potential to strengthen plants, improve commercial standards, enhance product quality, increase plant vitality, and facilitate harvesting (Kisvarga *et al*., 2022). Biostimulants, including protein hydrolysates and seaweed extracts, have been successfully utilized in horticultural crops promoting plant growth and improving crop quality (Paglialunga *et al*., 2022). These substances have been shown to promote the accumulation of carbohydrates and biosynthesis of compounds like anthocyanins in plants, contributing to improved fruit yield and quality (Lu *et al*., 2023).

Additionally, biostimulants have been found to increase mineral content, yield, and reduce nitrate content in horticultural crops, demonstrating their potential to enhance plant nutrition while reducing reliance on traditional fertilizers (El-Nakhel *et al*., 2023). Furthermore, the application of biostimulants derived from seaweed and yeast extracts has been shown to improve fruit development and quality in tomato plants, highlighting their positive impact on crop production (Mannino *et al*., 2020). Furthermore, microbial biostimulants have been proven to enhance plant growth, increase nutrient availability, improve soil quality, and boost plants' tolerance to abiotic stress in various horticultural crops (Shahrajabian *et al*., 2023). Plant biostimulants have been shown to positively influence soil health by modulating soil biological and physicochemical properties (Koleška *et al*., 2017, Shayen, 2023, Wadduwage, 2023). By leveraging the benefits of biostimulants derived from natural sources and microorganisms, horticultural practices can optimize plant performance and contribute to sustainable agricultural production. Despite the active role of soil microorganisms in organic matter decomposition, nutrient cycling, and crop production, the mechanisms behind their activity remain largely unknown. Moreover, the effects of their application are often inconsistent, highlighting the need for further research to develop more precisely targeted products (Fadiji *et al*., 2022). Research on shifts in microbial community structure following the use of various biostimulants can facilitate the development of best management practices for agroecosystems.

The aim of the current study was to evaluate the metabolic activity and functional indices in soil treated with two different types of biofertilizers: seaweed-based and microbial, by taking samples at times when a greater effect on soil microorganisms was expected.

Materials and methods

Greenhouse experiment

The experiment was carried out in the private greenhouse located near Asenovgrad, Plovdiv Municipality. The greenhouse measured 13.50 m in length, 2.60 in width and 3 m in height (latitude: 42.013371, longitude: 24.874939). The two biofertilizers were compared to the untreated control in a completely randomized design with three replicates of five plants per variant. The space between rows were 70 cm and between plants in the row – 40 cm. Soil tillage included deep plowing and harrowing. The previous crop was spinach. The tomato seedlings (*Solanum lycopersicum*) var. *Kalina* F1 was transferred in the greenhouse in March 2024.

Biostimulants

The soil-applied biostimulants used in the current study are available on the Bulgarian market with the commercial name *RaisaMix* (*Daymsa*, Spain) and *AzoFixPlus* (*Bioenergy* LT, Lithuania), respectively. According to the manufacturer, *RaisaMix* is defined as a root's enhancer in the group of physioStimulants with a main content of free amino acids – 10.7% w/w (12.0% w/v) and total nitrogen (N) 4.0% w/w (4.5% w/v). *RaisaMix* contains also oligopeptides, polypeptides, alginates, mannitol, oligosaccharides and polysaccharides, natural growth hormones, betaines, polyamines and vitamins which are derived from special high-quality extract of the sea plant *Ascophyllum nodosum*.

AzoFixPlus contains the free-living diazotroph *Paenibacillus polymyxa* MVY-024 at 1.2×10^{12} CFU/l concentration, B-group vitamins (B1, B3, B6), macroelements such as potassium (7.14 g/l), sodium (1.88 g/l), calcium (1.50 g/l), sulfur (1.17 g/l), phosphorus (0.278 g/l), magnesium (0.275 g/l), and microelements (Cu, Co, Fe, Mn, Mo, Zn) with a maximum quantity of 0.02%.

In this study, both biostimulants were applied at a rate of 3 L/ha, following the manufacturers' recommendations. To achieve the prescribed dosage, 2 mL of the original stock solution was diluted in 10 mL of tap water and administered at 0.5 L per plant directly into the root zone. Applications were carried out twice, at ten-day intervals, during growth stages 19–22 as defined by the BBCH scale (Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie). For clarity, the treatment variants are referred to as follows: RaisaMix is abbreviated as PStim, and AzoFixPlus as MbPB.

Soil sampling

Soil samples for both chemical and microbiological analyses were taken after 60 days of the first application of biostimulants. Soil was collected from the rhizosphere of two uprooted plants and pooled to generate a composite sample for each treatment. This approach was intended to reduce the influence of local spatial heterogeneity within the rhizosphere and to obtain a representative sample of each treatment. However, sample pooling may reduce the ability to assess plant-to-plant variability and could lead to an underestimation of within-treatment variability. The soil was cleaned from stones, debris and biological materials, air-dried at room temperature, sieved using a 2-mm sieve and further processed according to requirements of analyses.

Soil chemical analyses

Soil analysis was done at the Accredited Laboratory complex of the Agricultural University, Plovdiv, Bulgaria and here the corresponding analyses are presented briefly. Total mobile nitrogen was determined using the modified

Kjeldahl method. Soil pH was determined potentiometrically in a soil-water solution. Soil phosphorus analysis was conducted according to the Egner *et al*. (1960) method. The applied method, with slight modification, includes $SnCl_2$ as an indicator and measurement of wavelength at 700 nm (BG GOST 26209-91). Available potassium was determined with a flame photometer (GOST 26209:1992).

Soil respiration

Soil respiration was determined according to Isermeyer method (1952) as described in Alef & Nannipieri (1995). The method is based on soil respiration in closed jars and the quantity of CO_2 absorbed by NaOH solution and consequently titrated with HCl. The following formula is used:

$$CO_2(mg)/SW/t = \frac{(V_0 - V)\times 1.1}{dwt},$$

where SW – the amount of soil dry weight in g, t is incubation time in hours, V_0 is quantity (in ml) of HCl for blanks sample titration (average), V is quantity (in ml) of HCl for soil sample titration, dwt is the dry weight of 1 g moist soil, 1.1 is the conversion factor (1 ml 0.05 M NaOH equals 1.1 mg CO_2).

Biolog EcoPlate technique

Biolog EcoPlate™ (Biolog Inc., USA) was used for assessment of metabolic activity of microbial communities. Each EcoPlate is comprised of three replicates of 31 different substrates organized in the following guilds - carbohydrates, carboxylic acids, polymers, amino acids, amines and phenols (Table 1).

Table 1. List of the substrates in the Biolog EcoPlate

Guild	Substrates
Amino acids	L-Arginine
	L-Asparagine
	L-Phenylalanine
	L-Serine
	L-Threonine
	Glycyl-L-glutamic acid
Amines	Phenylethylamine
	Putrescine

Guild	Substrates
Carboxylic acids	Pyruvic acid methyl ester
	D-Glucosaminic acid
	D-Galacturonic acid
	γ-Amino butyric acid
	Itaconic acid
	α-keto Butyric acid
	D-Malic acid
Phenols	2-Hydroxybenzoic acid
	4-Hydroxybenzoic acid
Carbohydrates	D-Cellobiose
	α-D-Lactose
	β methyl D Glucoside
	D -Xylose
	i-Erythritol
	D-Mannitol
	N-acetyl-D-glucosamine
	Glucose-1-phosphate
	D, L-α-Glycerol phosphate
	D-Galactonic acid γ-lactone
Polymers	Tween 40
	Tween 80
	A-Cyclodextrin
	Glycogen

Ten grams of preliminary prepared soil was suspended in 90 ml sterile physiological solution (0.85% NaCl), thoroughly mixed for 30 min at 190 rpm and left to settle for 10 min, after which a 10^{-3} dilution was prepared. One hundred and fifty µl were used for the Biolog EcoPlates inoculation as recommended by the producer`s protocol. The plates were incubated at 24±1°C and read spectrophotometrically at 24-hour intervals for 7 days (168 hours) with the MicroStation™ Reader provided by the Biolog system. The calculations for average well-color development (AWCD) and separately for each substrate guild were based on the optical density (OD) measured at 590 nm and 750 nm according to the procedure described by Sofo & Ricciuti (2019), except the formula for AWCD which was according to Huang *et al*. (2012) as follows:

$$\text{AWCD} = \sum(C_i - R)/31,$$

where R is the control well (water) and C_i is the value of any substrate-containing well. The measurement at 24-hour was used for data normalization by subtracting each consecutive measurement with the corresponding values at 24-hour in order to avoid so called background noise according to Urakawa *et al*. (2013). The negative numbers obtained during normalization were set to zero (Garland, 1996).

Functional indices

According to recommendation of Sofo & Ricciuti (2019) functional indices should be calculated based on wells with an OD ≥0.250. Due to very high metabolic activity of microorganisms in collected soil samples the current study implemented a different approach and the indices were calculated using the measurements from 48th to 168th hour of EcoPlate incubation taking into account wells with OD ≥0.350. The indices calculation was done with the formulas listed in the Table 2.

Table 2. Formulas for functional indices calculation

Functional indices	Formula	References
Shannon-Wiener index, H'	$H' = -\sum p_i \times (ln\, p_i)$ where pi is Ci, divided by the sum of Ci, wells with value ≥ 0.350	Jurkšienė *et al*. (2020)
Pielou index, E	$E = \frac{H'}{ln\, S}$ where H' is Shannon-Wiener index S – number of wells with value ≥ 0.350	Pielou (1966), Jurkšienė *et al*. (2020)
Simpson index, D	$D = 1 - \sum P_i^2$ where Pi is Ci, divided by the sum of Ci, wells with value ≥ 0.350	Chen *et al*. (2020a)
Margalef index, d	$d = \frac{(S-1)}{lnN}$ where S – number of wells with value ≥ 0.350, N – number of substrates i.e. 31	Türkmen & Kazanci (2010)
McIntosh index, U	$U = \sqrt{\sum P_i^2}$ where Pi is Ci, divided by the sum of Ci wells with value ≥ 0.350	McIntosh (1967) Huang *et al*. (2012)

Functional indices	Formula	References
McIntosh evenness, McI	$McI = N - U/N - (N/\sqrt{S})$ where U – McIntosh diversity index, N – sum of wells with value ≥ 0.350, S – number of substrates i.e. 31	Xu *et al*. (2015)
Gini coefficient, G	$G = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\lvert x_i - x_j\rvert}{2N^2\bar{x}}$ where x_i and x_j represent each pair of OD readings, $\bar{x}$ - AWCD, N - number of substrates. The final value was further multiplied with n/(n-1)	Weiner & Solbrig (1984) Harch *et al*. (1997)

Data analysis

The average values of AWCD, expressed both as total activity and per guild, were visualized using Microsoft Excel. Each EcoPlate's three sets of substrates were treated as replicates (n=3). Error bars indicate the standard deviation. Functional indices were compared using ANOVA, followed by Tukey's Honest Significant Difference (HSD) post-hoc test in SPSS (IBM, version 26).

Results

Soil analysis

Soil pH, ammonium and nitrate content of the soil were not affected by application of biostimulants (Table 3). There was a slight increase in the phosphorus content of soil samples treated with biostimulants but the potassium content was the same in the control and treated with microbial biostimulant sample and slightly lower in the variant which was treated with a seaweed-based biostimulants.

Table 3. Soil chemical parameters after application of two different biostimulants

Soil parameter	Units	Variant		
		Control	PStim	MbPB
pH	-	8.15	8.30	8.26
Ammonium (NH_4^+)	mg/kg	7.64	8.34	7.00
Nitrate (NO_3^-)	mg/kg	15.93	18.24	14.59
Phosphorus (P_2O_5)	mg/100g	98.61	100.16	107.91
Potassium (K_2O)	mg/100g	93.48	86.62	93.87

Soil respiration

Soil respiration was higher for variants that were treated with biostimulants, with corresponding values of 0.174 ± 0.010 and 0.162 ±0.018 CO_2(mg)/SW/24h for MbPB and PStim treatment, respectively. The control variant had a soil respiration of 0.124 ±0.014 CO_2 (mg)/SW/24h (Figure 1).

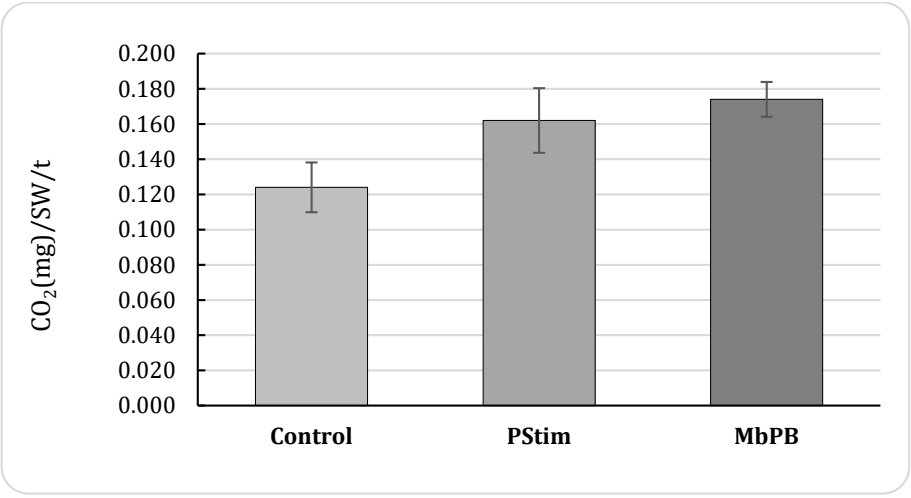


Figure 1. Soil respiration (PStim=RaisaMix; MbPB= AzoFixPlus).

Microbial metabolic activity

After the initial lag phase (0–24 hours), which is typical for microorganisms adapting to a new environment and newly available substrates, the microbial communities in all three tested samples exhibited a rapid and pronounced increase in metabolic activity, as measured by the Biolog EcoPlate (Figure 2). Between 48 and 120 hours, the increase in optical density (OD) was proportional, with a margin of approximately 0.2 units. Statistically, the mean values during this period showed no significant differences between the variants. However, after 96 hours, the control treatment displayed slightly higher values compared to the biostimulant-treated samples.

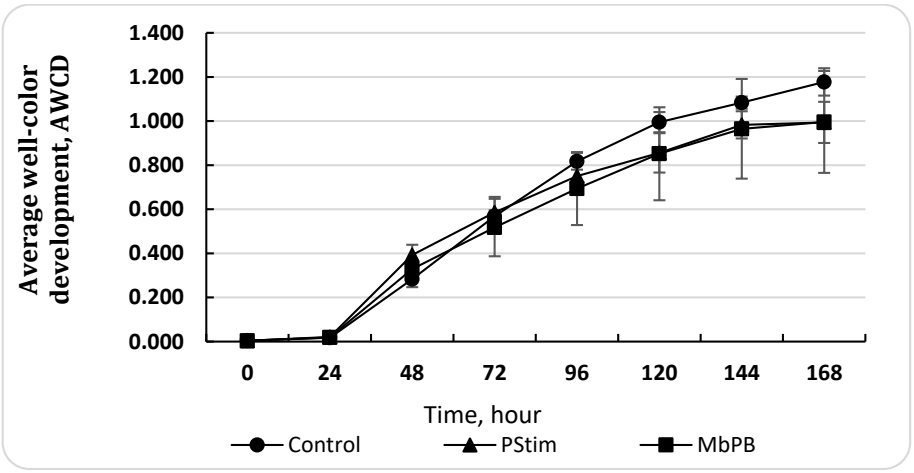


Figure 2. Overall metabolic activity (AWCD) of microbial communities in soil treated with two different biostimulants (PStim=RaisaMix; MbPB= AzoFixPlus).

Despite the lack of a significant difference in the metabolic activity expressed as AWCD between treatments, they revealed some differences in the utilization of different substrates. Utilization of amino acids after the 72nd hour until the end of the incubation period was higher in the control, and at the 96th hour, this difference was even statistically significant (Figure 3).

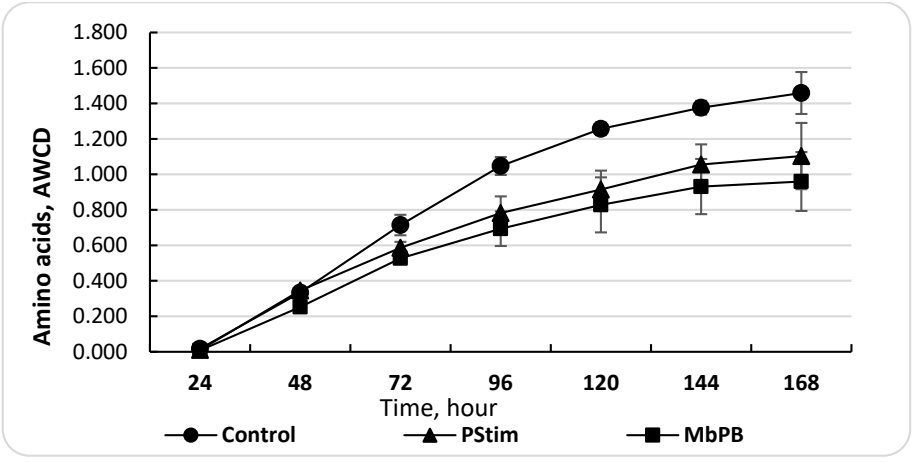


Figure 3. Amino acids utilization by microbial communities in soil treated with different biostimulants (PStim=RaisaMix; MbPB= AzoFixPlus).

Utilization of amines was relatively higher in all treatments, and at the end of the incubation period, the highest values were measured for the community in the sample treated with microbial-based biostimulant – 1.280 ±0.379, followed by the control and PStim with mean values of 1.159 and 1.009, respectively (Figure 4). A statistical difference was not obtained, probably because of the similar mean values and relatively wide range of standard deviation.

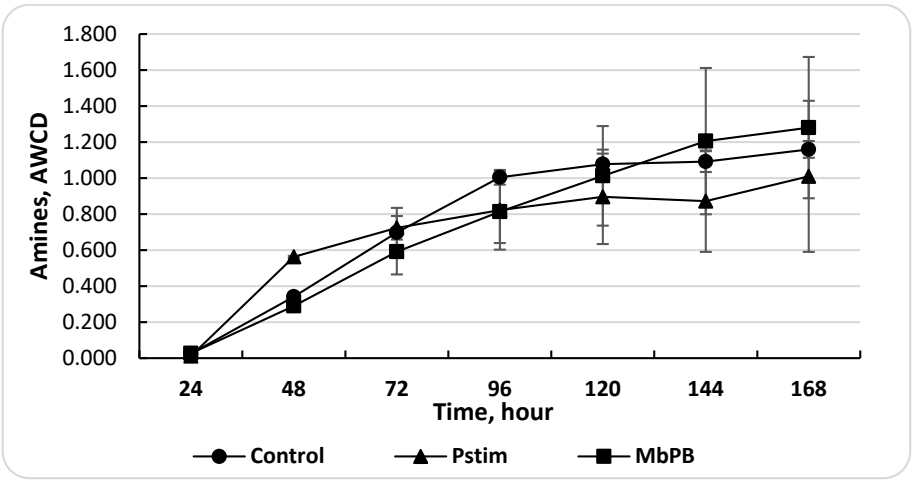


Figure 4. Amines utilization by microbial communities in soil treated with different biostimulants (PStim=RaisaMix; MbPB= AzoFixPlus).

Carboxylic acids were utilized quite uniformly across the different treatments up to the 48th hour. After this point, the curve became steep and then plateaued from the 120th hour onward, with the last two optical density readings in the control variant remaining nearly constant (Figure 5). Microbial communities collected from soil that received biostimulants treatment utilized carboxylic acids less intensely in comparison to the control (1.260 ±0.108 at the 168th hour) but very uniformly. A statistical difference between the control and biostimulants-treated variants was proven for measurements taken from the 46th to the 96th hour and at the end of the incubation period (p=0.006).

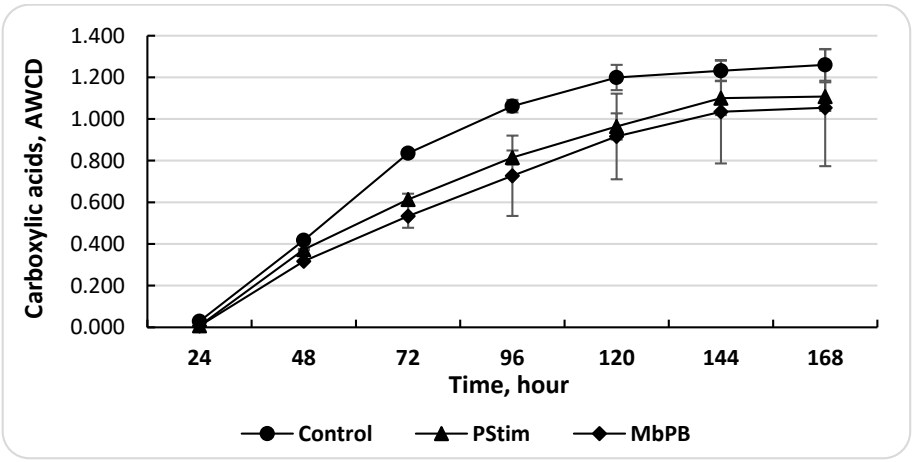


Figure 5. Carboxylic acids utilization by microbial communities in soil treated with different biostimulants (PStim=RaisaMix; MbPB= AzoFixPlus).

Contrary to the use of carboxylic acids, the communities treated with biostimulants showed significantly higher carbohydrate utilization (see Figure 6).

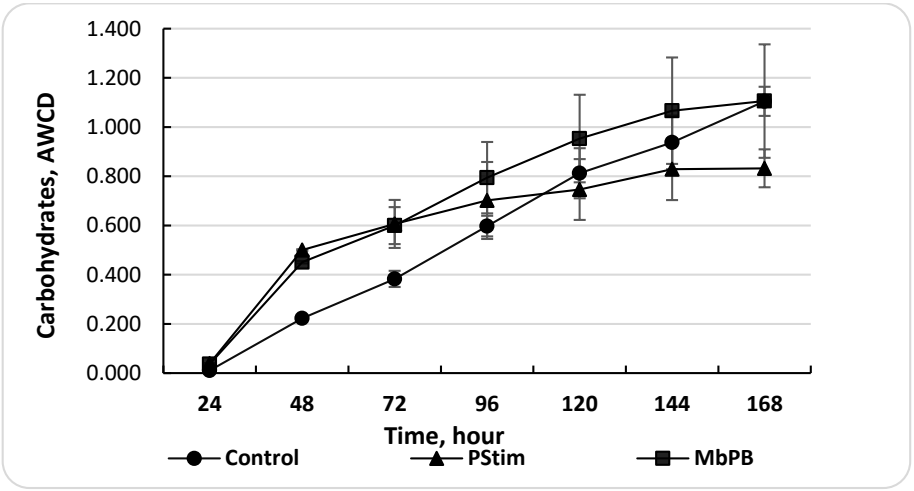


Figure 6. Carbohydrates utilization by microbial communities in soil treated with different biostimulants (PStim=RaisaMix; MbPB= AzoFixPlus).

By the 48th hour of incubation, the mean optical density (OD) was twice as high as that of the control treatment (0.222 ±0.033), and the difference was statistically significant. After the 96th hour, the communities treated with

seaweed-based biostimulants slowed their carbohydrate utilization, reaching the lowest value of 0.832 ± 0.08 by the end of the incubation period. Carbohydrate utilization in the control community was consistent and exponential throughout the incubation period. From the 96th to the 144th hour, the highest metabolic activity was observed in the community treated with microbial biostimulant. However, the final measurement showed that the mean values for the control and microbial biostimulant-treated samples were almost the same –1.105 ±0.130 and 1.106 ±0.240, respectively.

In comparison to the utilization of other available substrates in the Biolog EcoPlate, the polymers utilization was probably the most intriguing (Figure 7). Between the 24th and 48th hours, there was no difference between treatments. The community from the sample treated with microbial biostimulant showed almost the same OD mean values from the 48th to the 72nd hour, and after that, until the end of the incubation period, the increase did not exceed 0.1 units. As a result, the value at the final measurement was the lowest: – 0.497 ±0.124. In the current study, this was valid not only in comparison to the other experimental variants but also in comparison to the utilization of all other substrates' guild in the EcoPlate. The communities in the control variant and the treatment which received the seaweed-based biostimulant showed a stable and consistent increase in the OD, which reached almost 0.3 units from the 72nd to 96th hours, and then decreased to only 0.1-0.2 units. At the end of the incubation period, the variant treated with seaweed-based biostimulant had the highest mean value of 1.178 ±0.143, and the control reached 1.056 ±0.159.

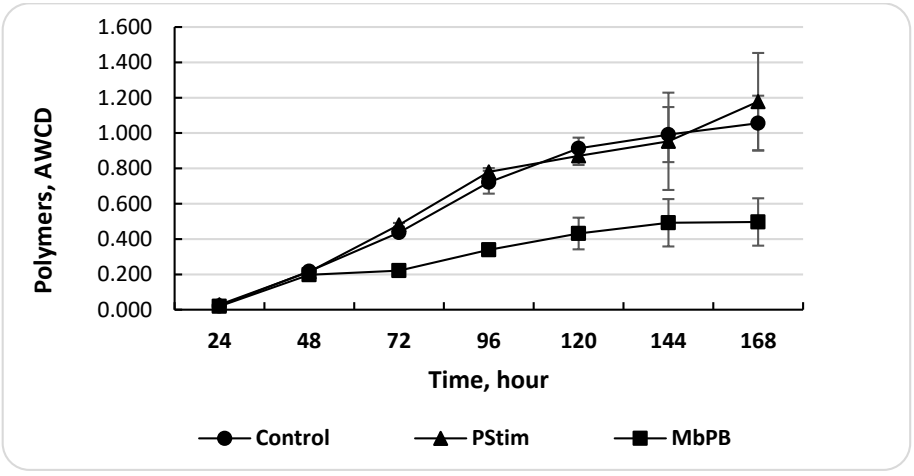


Figure 7. Polymers utilization by microbial communities in soil treated with different biostimulants (PStim=RaisaMix; MbPB= AzoFixPlus).

The metabolic activity of soil microbial communities towards phenolic compounds was primarily based on the utilization of 4-hydroxybenzoic acid, which greatly surpassed that of the second compound in the guild, 2-hydroxybenzoic acid, except in the variant treated with microbial biostimulant. Despite the highest utilization of phenols observed in the community collected from the MbPB variant, the uneven distribution of positive values in the wells containing 2-hydroxybenzoic acid caused a high standard deviation, as shown in the graph (Figure 8). As expected, at the end of the incubation period, the MbPB variant showed the highest utilization of phenols, reaching a value of 1.069 ±0.0558. This can be compared to its metabolic activity towards carboxylic acids and amino acids, with end values of 1.054 ±0.298 and 0.960 ±0.172, respectively. Microbial communities in samples that received no treatment and those treated with seaweed extract showed lower metabolic activity, with similar OD values at the end of the incubation period – 0.677 ±0.070 and 0.698 ±0.183, respectively.

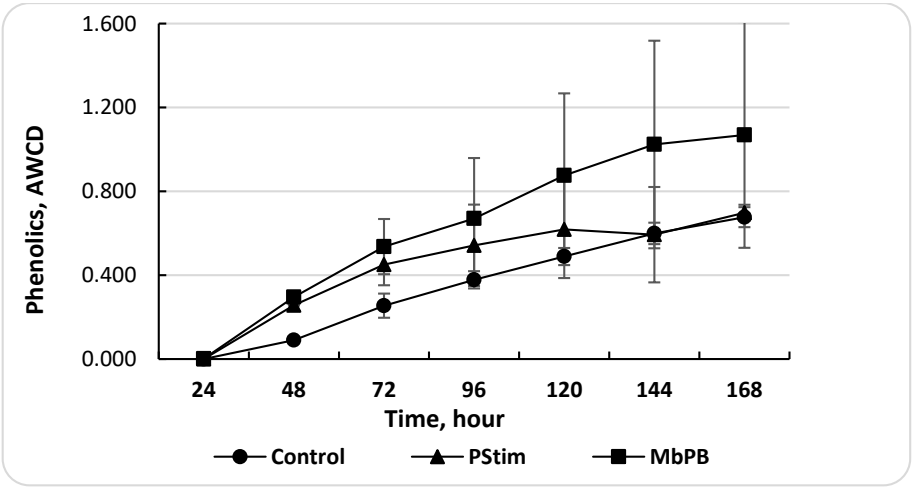


Figure 8. Phenols utilization by microbial communities in soil treated with different biostimulants (PStim=RaisaMix; MbPB= AzoFixPlus).

Functional indices

Despite the absence of statistically significant differences, some functional indices showed numerical increases in the biostimulant-treated samples, suggesting a potential trend toward enhanced metabolic diversity and evenness. However, indices such as Shannon-Weiner, Pielou, McIntosh evenness, and the Gini coefficient indicated higher functional diversity and evenness for variants treated with

seaweed-based and microbial biostimulants (Table 4). According to the Margalef index, higher biodiversity was observed in substrate utilization from sample treated with seaweed-based biostimulant, with the highest value of 6.49 ± 1.38. Conversely, the McIntosh index indicated higher metabolic diversity from soil treated with microbial biostimulant, with the highest value of 0.229 ±0.037.

Table 4. Functional indices of the characteristics of microbial communities in soil treated with two different biostimulants

Treatments	Functional indices						
	Shannon-Wiener index, H'	Pielou index, E	Simpson index, D	Margalef index, d	McIntosh index, U	McIntosh evenness, McI	Gini coefficient, G
Control	2.994 ±0.367	0.975 ±0.040	0.946 ±0.026	6.37 ±1.95	0.227 ±0.049	1.200 ±0.023	0.246 ±0.089
PStim	3.066 ±0.244	0.981 ±0.007	0.867 ±0.238	6.49 ±1.38	0.222 ±0.030	1.204 ±0.010	0.259 ±0.077
MbPB	3.015 ±0.291	0.984 ±0.005	0.946 ±0.019	6.16 ±1.58	0.229 ±0.037	1.201 ±0.015	0.256 ±0.083

Relative utilization of substrates

The relative utilization of substrates in the EcoPlate, based on measurements taken from the 48th to 168th hour of incubation, indicated higher utilization of amino acids and carboxylic acids at 22% and 21%, respectively, compared to the biostimulants-treated samples (Figure 9). The utilization of polymers in the control sample was comparable to those treated with the seaweed extract – 15% and 14%, respectively. However, the control sample showed significantly lower utilization of phenols – only 9% – compared to the samples treated with microbial biofertilizer (17%) and seaweed extract (13%). Both communities treated with either seaweed extract or microbial biostimulant showed similar metabolic activity towards utilization of amino acids (17% and 16%), amines (20%), carboxylic acids (18%), and carbohydrates (18% and 19%), respectively. The main difference between substrate utilization of communities in biostimulants-treated soil was utilization of polymers and phenolic compounds. In the former case, the variant that received the seaweed extract surpassed the sample with microbial biostimulant by 5%. Conversely, for phenolic compounds utilization, the sample with microbial biostimulant surpassed the sample that received the seaweed extract by 4%.

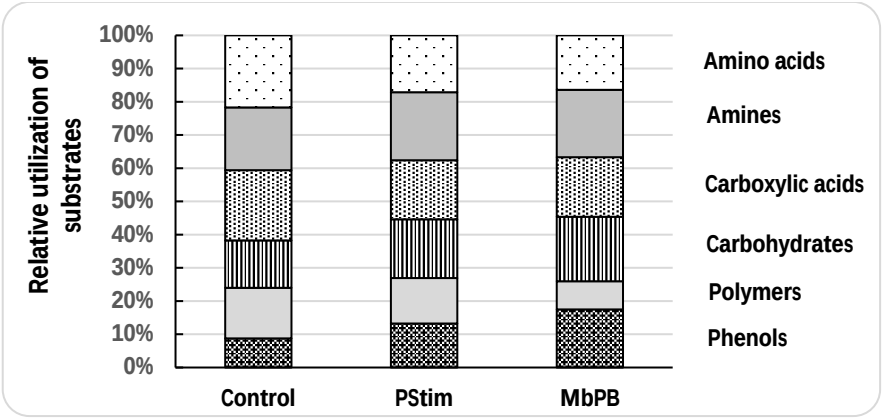


Figure 9. Relative utilization of substrates (%) in the Biolog Ecoplate by microbial communities in soil treated with two different biostimulants (PStim=RaisaMix; MbPB= AzoFixPlus)

Discussion

The application of biostimulants did not significantly affect the measured soil parameters. Similarly, none of the differences in the calculated functional indices characterizing substrate utilization pattern were statistically significant.

Although numerical differences were observed among treatments, these should be interpreted as trends rather than as evidence of treatment effects. The results should also be considered in light of certain limitations, including the controlled greenhouse conditions and the relatively small number of biological replicates, which may have limited the statistical power to detect subtle treatment effects. Despite the absence of significant changes in soil parameters under the conditions of the present study, contrasting results have been reported in the literature. For example, Ioppolo et al. (2020) observed an initial decrease of 2–3 pH units and higher electrical conductivity in samples treated with biofertilizer. However, the authors used in their study a citrus fruit processing wastewaters which supposedly could affect the aforementioned soil parameters. Chen et al. (2020b) also reported that application of seaweed-based extract increased the ammonium and nitrate content in comparison to the control. Conversely, Hellequin et al. (2018) reported a pH neutralizing effect of biostimulants application and increased organic carbon mineralization. The authors considered that these changes were linked to changes in microbial communities.

Currently, agricultural practices are evolving to achieve sustainable soil management, health, and productivity, while also supporting biodiversity through alternatives to common approaches. In this effort, agricultural biostimulants

are being developed and used as alternatives to mineral fertilizers. However, the ways in which these biostimulants enhance soil biological functions and indirectly boost crop yields remain unclear (Hellequin *et al*., 2018). The summarized net metabolic activity of microbial communities can be represented by changes in the OD throughout the incubation period as AWCD. In principle, in the absence of inhibitory supplements or other factors that can suppress microbial activity, the AWCD curve, similar to the observed in the current experiment, has a typical sigmoid shape (Stefanowicz, 2006; Lima *et al*., 2015). On the other hand, the length of the lag phase and the pace of the OD increase entirely depend on metabolic activity of community (Li *et al*., 2012; Koner *et al*., 2021). The very short lag phase observed in the current study is similar to the results of Ge *et al*. (2018) and contrasts with the study of Sun *et al*. (2012) who observed a lag phase of 24 and more than 60 hours, respectively. Despite the provision of summarized information, the AWCD cannot represent the specifics of substrate utilization. The calculation of AWCD separately per each of the six substrates guilds in the Biolog EcoPlate addresses this gap. Across the utilization of substrates, some specificities have been observed in the current study. In the case of utilization of amino acids and carboxylic acids the communities which were collected from samples treated with biostimulants firstly were quite similar and despite being considered as the main carbon and nitrogen sources they were metabolized less intensely in the biostimulants supplemented variants. This could be explained with changes in carbon and nitrogen microbial metabolic dynamics. This implication was made by Chen *et al*. (2002) who investigated the effects of two commercially available soil biostimulants through short-term (1 week) and longer-term (8 weeks) soil incubations under laboratory conditions. They found that in the short-term soil incubations, the two biostimulants had different effects on microbial activity and stimulated different properties and in the long-term experiment the biostimulants significantly affected soil nitrogen dynamics. The results of the study indicated that the two biostimulants could stimulate both the breakdown and mineralization of soil organic materials and the observed trend was related to the possibility of selective inhibition or stimulation of particular species in the microbial community (Chen *et al*., 2002). Ioppolo *et al*. (2020) suggested that the application of biostimulants changed the total and labile C pools, thus stimulate soil microbial activity and biomass. However, they assigned this effect predominantly on the transient soil acidification caused by the biofertilizer used. In similar cases of changes in soil pH after biofertilizer application the authors recommended soil electrical conductivity and soil pH monitoring.

Chen *et al* (2020b) studied the effect of seaweed fertilizer application and found that the overall functional pattern was similar between the different soil samples. However, the authors stressed on the fact that the relative abundance of functional proteins involved in carbohydrate transport and metabolism in the treatment groups have increased when compared to the control. This result corresponds with the higher utilization of carbohydrates which was observed in the current study in the biostimulants treated samples. Based on the observations and analysis Chen *et al*. (2020b) made the conclusion that the relative abundance of the microbiota varied significantly after biofertilizer application and there were also some changes in the enzyme activity.

Functional diversity constitutes a crucial aspect of overall metabolic diversity in soil, encompassing a wide range of activities. Biodiversity plays a vital role in maintaining ecosystem stability and resilience. However, due to the immense abundance and species diversity of microbial communities in soil, the precise relationship between diversity and metabolic activity remains largely unknown (Torsvik and Øvreås 2002). Li *et al.* (2022) conducted a meta-analysis on the effects of biofertilizers and concluded that the benefits of biostimulant application are greater under unfavorable soil and environmental conditions, particularly in non-neutral, saline, nutrient-poor sandy soils with low organic matter content. This suggests that the effects of biostimulant application may be less pronounced in fertile soils and under favorable growing conditions. Estimating bacterial functional diversity based on substrate utilization has been reported as a sensitive approach for detecting changes in soil conditions (Pessi *et al*., 2012; Zhang *et al*., 2013). Additionally, it allows for the comparison of microbial communities from different sources if certain considerations are taken into account (Preston-Mafham *et al*., 2002). Roesti *et al*. (2006) found that a biofertilizer application caused significant modifications in microbial community structure. In contrast, Baldi *et al*. (2021) and Wang *et al*. (2021) did not detect significant changes in soil microbial communities following biofertilizer application, or the observed effects were inconsistent. The results obtained by Wadduwage *et al*. (2023) also implied that biostimulant treatment did not significantly affect total bacterial and fungal abundance or their alpha diversity. Dal Cortivo *et al*. (2020) did not observe alterations in microbial biodiversity. Nevertheless, they argue that optimizing biofertilizer use in sustainable wheat cultivation necessitates attention to the composition of microbial consortia. Furthermore, Mickan *et al.* (2021) reported that microbial biostimulants influenced soil biodiversity but did not affect the growth of the plant species studied. Their results also revealed an interaction between plant type and soil amendments. The authors emphasized the complexity of interactions between soil amendments and microbial biostimulants, which can affect soil bacterial communities and warrant caution

when interpreting the relationship between soil biodiversity and plant growth. Hellequin *et al.* (2018), who conducted soil microcosm experiments with the addition of crop residues and a biostimulant, found that the increase in microbial biomass and crop residue mineralization was linked to changes in the soil microbial communities. However, due to the significant prevalence of decomposer species, the overall soil microbial richness and diversity actually decreased. Some authors have reported that soil application of microbial biostimulants increased bacterial diversity, whereas seed treatment reduced the diversity of indigenous plant-associated bacteria. The observed changes in bacterial diversity were attributed primarily to the response of r-strategist bacteria (Ciccillo *et al*., 2002). Similarly, other studies have shown that different fertilization practices increased the abundance of soil microorganisms while exerting only limited effects on microbial community structure. Nevertheless, these practices may influence the catabolic activity of fast-growing bacterial populations (Guanghua *et al*., 2008). The results of the present study suggest that the applied biostimulants may have selectively affected specific functional groups within the soil microbial community. However, because no taxonomic analyses were conducted, the possible predominance of fast-growing copiotrophic microorganisms remains speculative.

Conclusions

The growing use of biostimulants in agriculture is driven by sustainability goals and scientific research. As partial or full alternatives to chemical fertilizers, they support eco-friendly farming, though their effectiveness varies with environmental and biological conditions. Soil-applied biostimulants can alter metabolic activity and substrate utilization, influencing nutrient turnover and soil health. While no changes in chemical composition were observed in the current experiment, treated and untreated soils showed very high metabolic activity, limiting detection of subtle differences. Biostimulant-treated soils showed a tendency toward greater metabolic diversity and evenness in substrates utilization, although these differences were not statistically significant. These effects may stem from selective stimulation of specific microbial species, warranting further research into their role in soil fertility.

Competing interests. The author declares no competing interests

Acknowledgements. This study was financially supported by the Centre for Science Research, Technology Transfer, and Protection of Intellectual Property at the Agricultural University – Plovdiv, within the scope of Project 03/23.

References

Alef, K. & Nannipieri, P. (1995) Soil respiration. In: *Methods in applied soil microbiology and biochemistry*, Alef, K. & Nannipieri, P. (eds.), 114-116

Baldi, E., Gioacchini, P., Montecchio, D., Mocali, S., Antonielli, L., Masoero, G., & Toselli, M. (2021). Effect of biofertilizers application on soil biodiversity and litter degradation in a commercial apricot orchard. *Agronomy,* 11(6), 1116. https://doi.org/10.3390/agronomy11061116

Chen, S. K., Subler, S., & Edwards, C. A. (2002). Effects of agricultural biostimulants on soil microbial activity and nitrogen dynamics. *Applied Soil Ecology*, 19(3), 249-259. https://doi.org/10.1016/S0929-1393(02)00002-1

Chen, S. N., Shang, P. L., Kang, P. L., & Du, M. M. (2020a). Metabolic functional community diversity of associated bacteria during the degradation of phytoplankton from a drinking water reservoir. *International Journal of Environmental Research and Public Health*, 17(5), 1687. https://doi.org/10.3390/ijerph17051687

Chen, Y., Li, J., Huang, Z., Su, G., Li, X., Sun, Z., & Qin, Y. (2020b). Impact of short-term application of seaweed fertilizer on bacterial diversity and community structure, soil nitrogen contents, and plant growth in maize rhizosphere soil. *Folia Microbiologica*, 65(3), 591-603. https://doi.org/10.1007/s12223-019-00766-4

Ciccillo, F., Fiore, A., Bevivino, A., Dalmastri, C., Tabacchioni, S., & Chiarini, L. (2002). Effects of two different application methods of *Burkholderia ambifaria* MCI 7 on plant growth and rhizospheric bacterial diversity. *Environmental Microbiology*, 4(4), 238–245. https://doi.org/10.1046/j.1462-2920.2002.00291.x

Dal Cortivo, C., Ferrari, M., Visioli, G., Lauro, M., Fornasier, F., Barion, G., Panozzo, A., & Vamerali, T. (2020). Effects of seed-applied biofertilizers on rhizosphere biodiversity and growth of common wheat (*Triticum aestivum* L.) in the field. *Frontiers in Plant Science*, 11, 72. https://doi.org/10.3389/fpls.2020.00072

El-Nakhel, C.; Petropoulos, S.A.; Di Mola, I.; Ottaiano, L.; Cozzolino, E.; Rouphael, Y., & Mori, M. (2023). Biostimulants of different origins increase mineral content and yield of wild rocket while reducing nitrate content through successive harvests. *Horticulturae,* 9, 580. https://doi.org/10.3390/horticulturae9050580

EU Regulation 2019/1009 of the European Parliament and of the Council of 5 June 2019 laying down rules on the making available on the market of EU fertilizing products and amending Regulations (EC) No 1069/2009 and (EC) No 1107/2009. Official Journal of European Union*,* 170, 1–114.

Fadiji, A.E., Babalola, O. O., Santoyo, G., & Perazzolli, M. (2022). The potential role of microbial biostimulants in the amelioration of climate change-associated abiotic stresses on crops. *Frontiers in Microbiology*, 12, 829099. https://doi.org/10.3389/fmicb.2021.829099

FAO (2021). The share of food systems in total greenhouse gas emissions: Global, regional and country trends 1990–2019. Available at: https://www.fao.org/food-agriculture-statistics/data-release/data-release-detail/en/c/1454718/.

Garland, J. L. (1996). Analytical approaches to the characterization of samples of microbial communities using patterns of potential C source utilization. *Soil Biology and Biochemistry*, 28(2), 213-221. https://doi.org/10.1016/0038-0717(95)00112-3

Ge, Z., Du, H., Gao, Y., & Qiu, W. (2018). Analysis on metabolic functions of stored rice microbial communities by BIOLOG ECO microplates. *Frontiers in Microbiology,* 9, 1375. https://doi.org/10.3389/fmicb.2018.01375

Guanghua, W., Junjie, L., Xiaoning, Q., Jian, J., Yang, W., & Xiaobing, L. (2008). Effects of fertilization on bacterial communitystructure and function in a black soil of Dehui region estimated by Biolog and PCR-DGGE methods. *Acta Ecologica Sinica*, 28, 220–226. https://doi.org/10.1016/S1872-2032(08)60023-2

Harch, B. D., Correll, R. L., Meech, W., Kirkby, C. A., & Pankhurst, C. E. (1997) Using the Gini coefficient with BIOLOG substrate utilisation data to provide an alternative quantitative measure for comparing bacterial soil communities. *Journal of Microbiological Methods*, 30(1), 91-101. https://doi.org/10.1016/s0167-7012(97)00048-1

Hellequin, E., Monard, C., Quaiser, A., Henriot, M., Klarzynski, O., & Binet, F. (2018). Specific recruitment of soil bacteria and fungi decomposers following a biostimulant application increased crop residues mineralization. *PloS One*, 13(12), e0209089. https://doi.org/10.1371/journal.pone.0209089

Huang, H. Y., Zhou, P., Shi, W. W., Liu, Q. L., Wang, N., Feng, H. W., & Zhi, Y. E. (2012). Microbial functional diversity in facilities cultivation soils of nitrate accumulation. *Procedia Environmental Sciences*, 13, 1037–1044. https://doi.org/10.1016/j.proenv.2012.01.097

Ioppolo, A., Laudicina, V. A., Badalucco, L., Saiano, F., & Palazzolo, E. (2020). Wastewaters from citrus processing industry as natural biostimulants for soil microbial community. *Journal of Environmental Management*, 273, 111137. https://doi.org/10.1016/j.jenvman.2020.11113

Jurkšienė, G., Janušauskaitė, D., & Baliuckas, V. (2020). Microbial community analysis of native *Pinus sylvestris* L. and alien *Pinus mugo* L. on dune sands as determined by ecoplates. *Forests*, 11(11), 1202. http://dx.doi.org/10.3390/f11111202

Kisvarga, S., Hamar-Farkas, D., Boronkay, G., Neményi, A., & Orlóci, L. (2022). Effects of biostimulants in horticulture, with emphasis on ornamental plant production. *Agronomy*, 12(5), 1043. https://doi.org/10.3390/agronomy12051043

Koleška, I., Hasanagić, D., Todorović, V., Murtić, S., Klokić, I., Parađiković, N., … & Kukavica, B. (2017). Biostimulant prevents yield loss and reduces oxidative damage in tomato plants grown on reduced NPK nutrition. *Journal of Plant Interactions*, 12(1), 209-218. https://doi.org/10.1080/17429145.2017.1319503

Koner, S., Chen, J. S., Hsu, B. M., Tan, C. W., Fan, C. W., Chen, T. H., Hussain, B., & Nagarajan, V. (2021). Assessment of carbon substrate catabolism pattern and functional metabolic pathway for microbiota of limestone caves. *Microorganisms*, 9, 1789. https://doi.org/10.3390/microorganisms9081789

Li, J. H., Jiao, S. M., Gao, R. Q., & Bardgett, R. D. (2012). Differential effects of legume species on the recovery of soil microbial communities, and carbon and nitrogen contents, in abandoned fields of the Loess Plateau, China. *Environmental Management*, 50(6), 1193–1203. https://doi.org/10.1007/s00267-012-9958-7

Li, J., Van Gerrewey, T., & Geelen, D. (2022). A meta-analysis of biostimulant yield effectiveness in field trials. *Frontiers in Plant Science*, 13, 836702. https://doi.org/10.3389/fpls.2022.836702

Lima, A. B., Cannavan, F. D. S., Germano, M. G., Dini-Andreote, F., de Paula, Frabchini, J. C., Teixeira, W. G., & Tsai, S. M. (2015). Effects of vegetation and seasonality on bacterial communities in Amazonian dark earth and adjacent soils. *African Journal of Microbiology Research*, 9(40), 2119-2134. https://doi.org/10.5897/AJMR2015.7453

Lu, Y., Liang, D., Xia, H., Pang, Y., Xiao, Q., Huang, Y., … & Lv, X. (2023). Biostimulants promote the accumulation of carbohydrates and biosynthesis of anthocyanins in 'yinhongli' plum. *Frontiers in Plant Science*, 13. https://doi.org/10.3389 /fpls.2022.1074965

Mannino, G., Campobenedetto, C., Vigliante, I., Contartese, V., Gentile, C., & Bertea, C. (2020). The application of a plant biostimulant based on seaweed and yeast extract improved tomato fruit development and quality. *Biomolecules*, 10(12), 1662. https://doi.org/10.3390/biom10121662

McIntosh, R. P. (1967). An index of diversity and the relation of certain concepts to diversity. *Ecology*, 48(3), 392–404. https://doi.org/10.2307/1932674

Mickan, B. S., Alsharmani, A. R., Solaiman, Z. M., Leopold, M., & Abbott, L. K. (2021) Plant-dependent soil bacterial responses following amendment with a multispecies microbial biostimulant compared to rock mineral and chemical fertilizers. *Frontiers in Plant Science*, 11, 550169. https://doi.org/10.3389%2Ffpls.2020.550169

Mughunth, R. J., Velmurugan, S., Mohanalakshmi, M., & Vanitha, K. (2024). A review of seaweed extract's potential as a biostimulant to enhance growth and mitigate stress in horticulture crops. *Scientia Horticulturae*, 334, 113312. https://doi.org/10.1016/j.scienta.2024.113312

Paglialunga, G., Proietti, S., Cardarelli, M., Moscatello, S., Colla, G., & Battistelli, A. (2022) Chicory taproot production: effects of biostimulants under partial or full controlled environmental conditions. *Agronomy*, 12(11), 2816. https://doi.org/10.3390/agronomy12112816

Pe'er, G., Bonn, A., Bruelheide, H., Dieker, P., Eisenhauer, N., Feindt, P. H., ... & Lakner, S. (2020). Action needed for the EU Common Agricultural Policy to address sustainability challenges. *People and Nature*, 2(2), 305-316. https://doi.org/10.1002/pan3.10080

Pessi, I. S., Elias, S. de O., Simoes, F. L., Simoes, J. C., & Macedo, A. J. (2012). Functional diversity of microbial communities in soils in the vicinity of Wanda Glacier, Antarctic peninsula. *Microbes and Environments*, 27(2), 200-203. https://doi.org/10.1264/jsme2.me11311

Pichyangkura, R., & Chadchawan, S. (2015). Biostimulant activity of chitosan in horticulture. *Scientia Horticulturae*, 196, 49-65. https://doi.org/10.1016/j.scienta.2015.09.031

Pielou, E. C. (1966). The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology*, 13, 131-144. https://doi.org/10.1016/0022-5193(66)90013-0

Preston-Mafham, J., Boddy, L., & Randerson, P. F. (2002). Analysis of microbial community functional diversity using sole-carbon-source utilisation profiles - a critique. *FEMS Microbiology Ecology*, 42(1), 1–14. https://doi.org/10.1016/S0168-6496(02)00324-0

Rathor, P., Gorim, L. Y., & Thilakarathna, M. S. (2023). Plant physiological and molecular responses triggered by humic based biostimulants - A way forward to sustainable agriculture. *Plant Soil,* 492, 31-60. https://doi.org/10.1007/s11104-023-06156-7

Roesti, D., Gaur, R., Johri, B., Imfeld, G., Sharma, S., Kawaljeet, K., & Aragno, M. (2006). Plant growth stage, fertiliser management and bio-inoculation of arbuscular mycorrhizal fungi and plant growth promoting rhizobacteria affect the rhizobacterial community structure in rain-fed wheat fields. *Soil Biology and Biochemistry*, 38(5), 1111–1120. https://doi.org/10.1016/j.soilbio.2005.09.010

Shahrajabian, M., Petropoulos, S., & Sun, W. (2023). Survey of the influences of microbial biostimulants on horticultural crops: case studies and successful paradigms. *Horticulturae*, 9(2), 193. https://doi.org/10.3390/horticulturae9020193

Shayen, M. (2023) An overview: seaweed extract as biostimulant. *International Journal of Progressive Sciences and Technologies*, 39(2), 304. https://doi.org/10.52155/ijpsat.v39.2.5491

Shelef, O., Fernández-Bayo, J. D., Sher, Y., Ancona, V., Slinn, H., & Achmon, Y. (2018). Elucidating local food production to identify the principles and challenges of sustainable agriculture. In: Sustainable food systems from agriculture to industry, Galanakis, C. M. (Ed.), 47-81. https://doi.org/10.1016/B978-0-12-811935-8.00002-0

Sofo, A., & Ricciuti, P. A. (2019). Standardized method for estimating the functional diversity of soil bacterial community by Biolog® EcoPlates™ assay - the case study of a sustainable olive orchard. *Applied Sciences*, 9(19), 4035. https://doi.org/10.3390/app9194035

Stefanowicz, A. (2006). The Biolog plates technique as a tool in ecological studies of microbial communities. *Polish Journal of Environmental Studies*, 15(5), 669-676.

Sun, Y. H., Yang, Z. H., Zhao, J. J., & Li, Q. (2012). Functional diversity of microbial communities in sludge-amended soils. *Physics Procedia*, 33, 726–731. https://doi.org/10.1016/j.phpro.2012.05.127

Tejada, M., Benítez, C., Gómez, I., & Parrado, J. (2011). Use of biostimulants on soil restoration: Effects on soil biochemical properties and microbial community. *Applied Soil Ecology*, 49, 11-17. https://doi.org/10.1016/j.apsoil.2011.07.009

Tiwari, S., Prasad, V., & Lata, C. (2019). *Bacillus*: Plant growth promoting bacteria for sustainable agriculture and environment. In: New and future developments in microbial biotechnology and bioengineering, Singh, J. S., Singh, D. P. (Eds), Elsevier, 43-55. https://doi.org/10.1016/B978-0-444-64191-5.00003-1

Torsvik, V., & Øvreås, L. (2002). Microbial diversity and function in soil: from genes to ecosystems. *Current Opinion in Microbiology*, 5, 240–245. https://doi.org/10.1016/S1369-5274(02)00324-7

Türkmen, G., & Kazanci, N. (2010). Applications of various biodiversity indices to benthic macroinvertebrate assemblages in streams of a national park in Turkey. *Review of Hydrobiology*, 3(2), 111-125.

Urakawa, H., Ali, J., Ketover, R. D. J., Talmage, S. D., Garcia, J. C., Campbell, I. S., & Loh, A. N., Parsons, M. L. (2013). Shifts of bacterioplankton metabolic profiles along the salinity gradient in a subtropical estuary. *ISRN Oceanography*, ID 410814. http://dx.doi.org/10.5402/2013/410814

Verfuerth, C., Bellamy, A. S., Adlerova, B., & Dutton A. (2023) Building relationships back into the food system: addressing food insecurity and food well-being. *Frontiers in Sustainable Food Systems*, 7. https://doi.org/10.3389/fsufs.2023.1218299

Wadduwage, J., Liu, H., Egidi, E., Singh, B. K., & Macdonald, C. A. (2023). Effects of biostimulant application on soil biological and physicochemical properties: A field study. *Journal of Sustainable Agriculture and Environment*, 2(3), 285-300. https://doi.org/10.1002/sae2.12057

Wang, J., Liu, L., Gao, X., Hao, J., & Wang, M. (2021). Elucidating the effect of biofertilizers on bacterial diversity in maize rhizosphere soil. *PLoS One*, 16(4). http://doi.org/10.1371/journal.pone.0249834

Weiner, J., & Solbrig, O. T. (1984). The meaning and measurement of size hierarchies in plant populations. *Oecologia*, 61, 334-336.

Xu, W., Ge, Z., & Poudel, D. R. (2015). Application and optimization of Biolog EcoPlates in functional diversity studies of soil microbial communities. *MATEC Web of Conferences*, 22, 04015. https://doi.org/10.1051/matecconf/20152204015

Zhang, J., & Drury, M. (2024). Sustainable agriculture in the EU and China: A comparative critical policy analysis approach. *Environmental Science and Policy*, 157, 103789. https://doi.org/10.1016/j.envsci.2024.103789

Zhang, N., Liu, W., Yang, H., Yu, X., Gutknecht, J. L. M., Zhang, Z., … & Ma, K. (2013). Soil microbial responses to warming and increased precipitation and their implications for ecosystem C cycling. *Oecologia*, 173(3), 1125–1142. https://doi.org/10.1007/s00442-013-2685-9.

STUDIA UNIVERSITATIS BABEŞ-BOLYAI BIOLOGIA, 71, 1, 2026 (p. 185-218)
doi:10.24193/subbbiol.2026.1.09

Diatom communities in freshwater springs with contrasting geology in the Apuseni Mountains

Anca-Mihaela Şuteu[1,2,3,4]✉, Laura Momeu[†,3], and Mihai Puşcaş[1,2,3,4]

[1]*Doctoral School of Integrative Biology, "Babeş-Bolyai" University, Cluj-Napoca, Romania;*
[2]*A. Borza Botanical Garden, "Babeş-Bolyai" University, Cluj-Napoca, Romania;*
[3]*Faculty of Biology and Geology, "Babeş-Bolyai" University, Cluj-Napoca, Romania;*
[4]*Center for Systematic Biology, Biodiversity and Bioresources - 3B,*
Faculty of Biology and Geology, Babeş-Bolyai University, Cluj-Napoca, Romania
✉***Corresponding author, E-mail: anca.ciorca@ubbcluj.ro.***

Article history: Received 13 March 2026; Revised 20 May 2026;
Accepted 20 May 2026; Available online 25 June 2026

Abstract. Freshwater springs are unique and fragile aquatic ecosystems characterized by a mosaic of microhabitats, hydrochemical properties influenced by geological substrate, and specialized biological communities. Springs from the Romanian Carpathian Mountains are poorly explored and the role of geological substrate in structuring diatom assemblages remains unclear. This study aims to characterize diatom communities and their dynamics in ten springs developed on contrasting geological substrates. Epibryon and epilithon samples were collected from at least three points of available microhabitats within the spring area. Standard methods were applied to obtain cleaned diatom material and the resulting assemblages were analyzed using multivariate ordination techniques, alongside ecological indicator values and functional guild structure. A total of 126 diatom taxa were identified and three of the investigated springs are reported here for the first time with respect to their diatom flora. Community composition differed significantly between the two contrasting groups of springs. Crystalline–metamorphic springs were associated with assemblages dominated by taxa typical of acidic and weakly

mineralized waters, whereas carbonate springs supported communities characteristic of more alkaline conditions. Low-profile guilds dominated carbonate springs, whereas one crystalline–metamorphic spring showed a distinct high-profile-dominated structure. The results indicate that geological substrate shapes diatom community dynamics primarily through its control of physicochemical water properties. Indicator species analysis (IndVal) identified *Pinnularia subcapitata* W. Gregory as associated with crystalline–metamorphic springs, whereas *Cocconeis euglypta* Ehrenberg, *Cocconeis placentula* Ehrenberg and *Meridion circulare* C.Agardh were linked to carbonate-dominated springs. This study provides one of the first integrative multivariate assessments of crenic diatom communities in the Apuseni Mountains and highlights contrasting ecological patterns associated with different geological settings.

Keywords: Carpathian Mountains, carbonate substrate, metamorphic schist, diatom guilds.

Introduction

Spring ecosystems are characterized by low hydrological variability, hydrochemistry largely determined by the geological substrate, and pronounced microhabitat diversity (Cantonati and Pipp, 2000; Fránková *et al*., 2009). In mountainous regions, particularly where springs emerge from carbonate or crystalline bedrock, distinct hydrochemical signatures are reflected in parameters such as electrical conductivity and buffering capacity (Ford and Williams, 2007). These environmental properties create crenic refuge habitats that support specialized and often stenotopic biotic communities (Taxböck *et al.*, 2017). The need to study and protect springs is emphasized by their small size and high biotic diversity, as any local disturbance can be detrimental (Fernández-Martínez *et al.*, 2024).

Diatoms (phylum Bacillariophyta) are globally distributed microalgae that occur across most aquatic habitats worldwide, with many taxa exhibiting cosmopolitan distributions (Vanormelingen *et al.*, 2008; Malviya *et al.*, 2016). However, numerous species exhibit narrow ecological tolerances and clear separation along environmental gradients such as conductivity, pH, nutrients or discharge variability (Cantonati *et al.*, 2022; Blanco *et al.*, 2025). Consequently, diatoms are widely used as bioindicators of the ecological status of aquatic ecosystems (Bellinger and Sigee, 2010). In springs, epibryon (epiphytic diatoms living on the surface of bryophytes) and epilithon (taxa growing on the surface of stones) represent specific microhabitat associated communities. Bryophytes

provide a complex structural substrate to grow on, with enhanced moisture buffering and microhabitat heterogeneity (Letáková *et al.*, 2018). Often, they host assemblages that differ in composition and diversity from epilithon (Cantonati *et al*., 2012a). In eucrenal habitats, these diatom communities reflect the distinctive hydrological characteristics of groundwater-fed ecosystems (Cantonati *et al*., 2012b).

Geological differences, reflected by distinct physical and chemical features, directly influence the diatom assemblages in aquatic habitats (Cantonati 1998; Cantonati *et al*., 2012a). Shifts in the benthic diatom communities can be observed when hydrochemical factors change the ionic composition or pH stability (Soininen 2007; Fránková *et al*., 2009). Carbonate spring systems typically exhibit higher alkalinity, greater buffering capacity, and elevated conductivity (Cantonati *et al.*, 2012b) compared to springs emerging from crystalline or siliceous bedrock (Beauger *et al.*, 2023).

Extensive research on crenic diatom communities in the European Alps has employed multivariate ecological modeling to better understand the structure of spring communities (Cantonati *et al*., 2006; Cantonati *et al*., 2012b). These studies emphasize the influence of bedrock lithology, discharge regime, and hydrochemical buffering capacity in shaping benthic communities, as demonstrated in karst systems of the Alps and the Mediterranean region (Lai *et al*., 2020), as well as mineral springs complexes such as Châteldon, Vichy and the Saint-Yorre basins (Baker *et al*., 2022). Springs are widely recognized as biodiversity hotspots, dependent on continuous input and highly sensitive to climatic and anthropogenic pressures (Angeli *et al*., 2010; Cantonati *et al*., 2012a).

In contrast, in the Carpathian Mountains, studies on spring diatoms have largely focused on documenting species composition and exploring ecological gradients in karst and crystalline environments (Fránková et al., 2009; Pouličková et al., 2005; Micle et al., 2018; Şuteu et al., 2025). Several studies have explored spring typology, hydrological and chemical variability, and crenic phytobenthos structure, contributing with valuable floristic and ecological knowledge from the several mountain regions like the Cracow–Częstochowa Upland and the Western Carpathians (Wojtal and Sobczyk, 2006; Fránková *et al*., 2009; Hájková *et al*., 2011; Wojtal and Sobczyk, 2012) and only more recently Romanian Carpathians, specifically the Apuseni Mountains (Micle *et al*., 2018; Şuteu *et al*., 2025). However, comprehensive analyses addressing both geological and environmental factors in spring habitats, while addressing functional bioindicator systems, remain limited in this region.

The Apuseni Mountains provide an ideal setting for disentangling these relationships, as contrasting lithological units compose this geographic area. The present study aims to clarify these relationships through an integrated

approach that combines community composition analysis with multivariate modelling across contrasting geological substrates. Based on previous research, we hypothesized that geological substrate would explain a significant proportion of variation in diatom community composition due to contrasting hydrochemical profiles. Therefore, we analyzed the crenic diatom communities found in ten springs located in the Apuseni Mountains (Romania) to determine the relative contribution of geology and physicochemical parameters in structuring them. Specifically, we (i) describe community composition and alpha diversity patterns across springs, (ii) assess the extent of community differentiation among geological substrates and (iii) quantify the influence of environmental predictors using multivariate analyses.

Materials and methods

Study area

The sampling campaign included ten freshwater springs located in north-western Romania, within the Apuseni Mountains (Fig. 1). The sampling sites span the Gilău–Muntele Mare Mountain group, the Pădurea Craiului Mountains and the central Apuseni Mountains karst area Padiș–Scărișoara from Bihorului Mountains. This spatial distribution encompasses both karstic and non-karstic hydrogeological settings, allowing comparisons across contrasting geological conditions (Orășeanu, 2016). Hydrologically, the studied springs are associated with three major river basins: the Someșul Rece, the Arieșul Mare (via one of its main tributaries, the Gârda Seacă), and the Crișul Repede (via one of its main tributaries, the Bucea river) (Orășeanu, 2020).

The investigated sites (Tab. 1) represent a range of spring types, from small headwater springs to large karst resurgences, characteristic of the Apuseni Mountain region. Four of the sampled springs are located within the boundaries of Parcul Natural Apuseni, Romania's largest protected karst area: GR4 to GR7. Springs located in the Gilău–Muntele Mare mountain group (MR1, MR2 and MR3) were sampled during July 2017, whereas the sampling of the remaining springs took place in November 2017. Three springs are sampled for the first time regarding their diatom flora: BC8, BC9 and SR10.

The spatial distribution of the investigated springs was mapped in the software R (version 4.5.2) using the packages sf and ggplot2 (Pebesma, 2018; Wickham, 2016). Field obtained spring coordinates were entered as point data in the WGS84 geographic coordinate system and then projected to UTM Zone 34N (EPSG:32634) for map visualization. A base map of Romania was obtained from the Natural Earth dataset (data source Natural Earth, 2024), and spring locations were plotted and symbolized according to geological group.

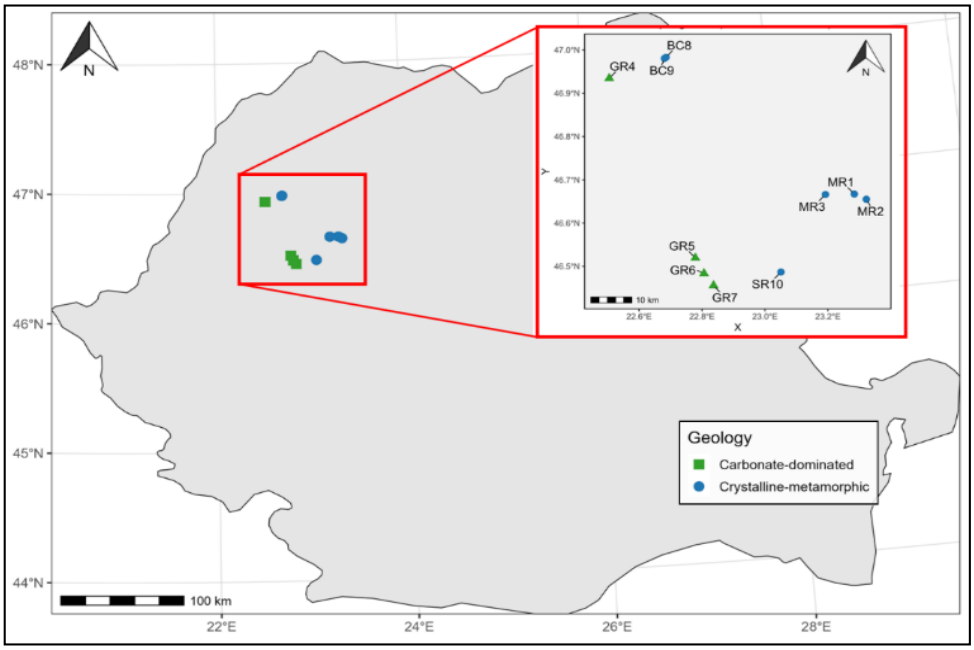


Figure 1. Geographic distribution of the investigated springs in the Apuseni Mountains, Romanian Carpathians (symbols indicate the geological substrate for each site; spring codes correspond to those listed in Tab. 1)

Geological and environmental classifications and assessment

Two geological categories were established based on lithological information obtained from the national geological map of Romania (1:200,000 scale), accessed via the INSPIRE geoportal of the Geological Institute of Romania (data source Geological Institute of Romania, 2018). Geological assignment was verified in QGIS (version 3.44.7) by overlaying spring coordinates onto this layer. For each spring, the lithological unit intersecting the sampling point was identified and used to assign substrate type. Two major geological groups were defined: carbonate-dominated springs, defined as draining limestone and dolomitic formations, characterized by carbonate bedrock and typically associated with higher buffering capacity and mineral content: GR4, GR5, GR6 and GR7; and crystalline–metamorphic springs, defined as draining metamorphic basement rocks (e.g., schists), generally associated with lower mineralization and reduced buffering capacity. The latter group comprised both: crystalline metamorphic (MR1, MR2, MR3, BC8 and BC9) and crystalline - siliceous metasedimentary bedrock (SR10). The applied classification reflects dominant lithology and associated hydrochemical regimes.

Springs were classified according to the crenic typology proposed by Cantonati *et al*. (2007), in: helocrene (the spring rivulet starts where the water emerging diffusely in the seepage area, begin to merge into a small channel), limnocrene (usually a small rim delimiting the spring pool as eucrenal area and the spring stream starts only below this rim) and rheocrene (the spring emerges at the surface through one or more flowing channels). The delimitation of the eucrenal and hypocrenal zones followed the criteria of Cantonati *et al.* (2007), with the eucrenal defined as the area within three meters of the spring emergence point, and the hypocrenal as the area located at least six meters from the same point.

The percentage of shade given by the canopy above the eucrenal area of each site was estimated following the methodology in Şuteu *et al*. (2025). In order to simplify the input of this environmental driver, the percentage values of the shaded area were categorized into three groups: sunny (exposed springs, with up to 25% of shaded areas), half shaded (covered eucrenal area between 26% and 50%) and shaded (from 51 to 100% coverage of the canopy area) (Tab. 1).

Flow intensity was assessed at the spring level using a semi-quantitative, field-based classification. Because comparable quantitative measurements could not be collected at all springs under geomorphological constraints, flow intensity was instead classified into three categories based on visual assessment of water movement: low flow (springs with apparently still water, no visible turbulence of the eucrenal area), medium flow (slow to moderate apparent current velocity), high flow (medium streams with high velocity and high turbulence of the eucrenal area). This categorical variable was assessed by the same person at each sampling site and used to capture ecologically relevant differences in hydrological regime among springs (Tab. 1).

Physico-chemical parameters were measured in situ using portable multiparameter: for pH and conductivity Hanna HI98130 and Hanna HI98194, and YSI-52 for water temperature and dissolved oxygen.

Diatom sampling methodology

At each spring site, two microhabitats were targeted: epilithon (biofilm on submerged stones) and epibryon (diatoms associated with submerged bryophytes). In some springs, additional samples were taken from both the eucrenal and hypocrenal area to account for within-site spatial heterogeneity (Tab. 1). Sampling was restricted to well-developed, spatially extensive microhabitats. At each site, three to five independent subsamples were collected per microhabitat. This increased spatial representativeness and captured small-scale variability. For bryophytes, one tuft was collected per subsample. For epilithon, an equivalent

surface of approximately 10 cm^2 area was scraped from a single stone for each subsample. Sampling intensity reflected the local availability of bryophyte substrates. At GR6, two bryophyte taxa were present in sufficient quantity to allow separate sampling, whereas at the other springs diatoms were analyzed from a single bryophyte sample per site.

Table 1. Spring names, codes and associated geological and hydrological characteristic considered for the present study (Abbreviations: K – karst or carbonate-dominated geology; C – crystalline-metamorphic geology; L – limnocrene; R – rheocrene; Hl – helocrene; E – eucrenal area; H – hypocrenal area; Sp - species)

Spring name	Spring code	Sampled microhabitat	Altitude (m)	Mountain group	Geology code	Spring type	Flow class	Shading
Muntele Rece	MR1	Bryophytes (E)	851	Gilău-Muntele Mare	C	L	Low	sunny
Muntele Rece	MR2	Bryophytes (E)	1036	Gilău-Muntele Mare	C	R	Low	shaded
Măguri Răcătău	MR3	Bryophytes (E)	626	Gilău-Muntele Mare	C	R	Low	sunny
Izbucul Izbândişului	GR4	Bryophytes (E, H)	406	Pădurea Craiului	K	L	High	half shaded
Izbucul Tăuzului	GR5	Bryophytes (E) Stones (E, H)	909	Bihorului	K	L	High	half shaded
Izbucul Coteţul Dobreştilor	GR6	Bryophytes (E) Sp1, Sp2 Stones (E, H)	807	Bihorului	K	R	Medium	shaded
Poarta lui Ionele	GR7	Bryophytes (E) Stones (E)	857	Bihorului	K	R	Low	shaded
Bătrânu	BC8	Bryophytes (E)	743	Pădurea Craiului	C	R	Medium	half shaded
Bucea	BC9	Bryophytes (E)	681	Pădurea Craiului	C	R	Low	half shaded
Somesul Rece	SR10	Bryophytes (E) Stones (E)	1521	Gilău-Muntele Mare	C	Hl	Low	sunny

The obtained subsamples from each spring were processed separately to isolate clean diatom frustules using a standard oxidation procedure, applied to both bryophyte and lithic material (Kelly *et al*., 1998). To ensure sufficient frustule abundance for reliable identification and counting, and to standardize analytical effort across sites, one composite sample was prepared separately for each microhabitat per spring by pooling the corresponding subsamples. Permanent slides were prepared by mounting two drops of homogenized suspension (approximately 0.1 mL) from a 10 mL diatom–water mixture. The identification of algal taxa was made to the species level using the following identification keys: Krammer and Lange-Bertalot (1986, 1988, 1991a, 1991b).

Three of the ten springs analyzed in this study (MR1, MR2, MR3) correspond to springs S1, S2, and S3 in Micle *et al*. (2018). For these springs, we used only the raw diatom and physico-chemical data, without reproducing any figures, tables, statistical results, or interpretations from the earlier publication.

Statistical analysis and applied indices

Taxonomic occurrence, defined as the number of taxa present in a sample, was organized in a matrix where 1 indicated the presence of a species in a sample and 0 indicated its absence. The frequency of the identified species was defined as the percentage from the total number of samples explained as: frequent species (a frequency above 80%, i.e., found in more than 8 springs), commonly occurring species (a frequency between 21% and 79%) and rare taxa (frequency of less or equal to 20%, i.e., present in one or two springs). These three categories were used as descriptive parameters throughout the text.

Composite samples representing different microhabitats were treated as separate ecological units because epilithon and epibryon support distinct diatom communities (Pouličková *et al*., 2004). Similarity among composite samples was calculated from a presence–absence matrix comprising all recorded diatom taxa, using the Jaccard similarity index in PAST software (version 4.14c; Hammer *et al*., 2001). In addition, similarity in quantitative community data was assessed using the Bray–Curtis index, computed from the abundance matrix of all the composite samples. Species richness (S), Shannon–Wiener diversity (H′), and Pielou's evenness (J) were calculated for each composite sample using PAST software. Differences in alpha diversity were explored descriptively across geological groups of springs. The Van Dam ecological indicator values (Van Dam *et al*., 1994) were calculated using OMNIDIA software (version 6.0.9; Lecointe *et al*., 1993), based on species relative abundances for each sample. The obtained results (salinity preference, oxygen requirements and moisture affinity) were used to support ecological interpretation of assemblage patterns. The following analysis was performed using the RStudio software (version 2026.01.0) and software R (version 4.5.2).

Indicator taxa associated with geological types and microhabitats were identified using the IndVal.g method (999 permutations) from indicspecies package in R software (De Cáceres and Legendre, 2009; De Cáceres, 2023). The ecological guild structure of diatom assemblages was assessed using the DiaThor package (Nicolosi Gelis *et* al., 2022; function *diat_guilds*). Taxa were assigned to ecological guilds and their relative contribution to each sample was calculated based on abundance data. Guilds are separated between: low-profile (small, adnate taxa closely attached to the substrate and tolerant to disturbance), high-profile (stalked or taxa forming colonies which extends into the water column, tolerating only low disturbance), motile (taxa capable of active movement across sediments), planktic (free-floating taxa typical of standing waters), and indeterminate. Guild composition was used to support ecological interpretation of community patterns (Rimet and Bouchez, 2012).

To visualize differences in diatom assemblage composition among composite samples from each spring, composite samples were projected on a non-metric multidimensional scaling (NMDS). Species abundances were first transformed to relative abundances and Bray–Curtis dissimilarity was calculated between composite samples. NMDS ordination was performed in two dimensions (k = 2) using the *metaMDS* function in the vegan package (Oksanen *et al*., 2024), with 200 random starts. Stress values were used to evaluate ordination fit. Associations between community composition and environmental variables were explored using the *envfit* function (999 permutations). Continuous variables (temperature, pH, dissolved oxygen, and conductivity), together with ordinal scores derived from canopy cover and flow category, were fitted separately from categorical factors (geology, canopy, and flow) onto the NMDS ordination. This approach allowed testing canopy cover and flow both as ordered gradients and as discrete categories, capturing potential directional trends as well as group-level differences in community composition.

Differences in community composition among geological types were tested using permutational multivariate analysis of variance (PERMANOVA; adonis2, 999 permutations), and homogeneity of multivariate dispersion was evaluated using *betadisper* function from vegan package.

Redundancy analysis (RDA) was conducted using the Hellinger-transformed community matrix to assess relationships between diatom assemblages and environmental predictors. Geological groups (carbonate-dominated vs crystalline–metamorphic) and measured environmental variables (conductivity, canopy cover, temperature, pH, dissolved oxygen) were included as explanatory variables. Categorical predictors were treated as factors. Predictors exhibiting no variation were excluded in the second RDA analysis, which is illustrated in this paper. Multicollinearity among predictors was evaluated using variance inflation factors (VIF), and variables with excessive collinearity were removed

where necessary. Model significance was assessed using permutation tests (999 permutations), including: overall model tests, marginal (unique) effects of predictors, axis-specific tests and adjusted R^2 values were calculated to estimate the proportion of explained variation.

To disentangle the relative contributions of geology and hydrochemical variables, variation partitioning (*varpart* function in vegan package) was performed using geology as one explanatory matrix and numeric environmental variables as a second one. Unique and shared fractions of explained variance were interpreted to assess the extent to which geological effects were mediated by environmental gradients.

Results

Qualitative floristic survey

Across the ten studied springs, the diatom community comprised a total of 126 taxa, of which: 54 were identified only in crystalline-metamorphic springs, 26 taxa only in karst springs and 46 were found in both spring groups (Supplementary Table 1). Within the diatom communities identified in this study, a total of 50 genera were recorded. The genera with the highest number of taxa were *Navicula* and *Pinnularia*. Taxon richness varied markedly among sites, ranging from a minimum of 13 taxa in crystalline springs to a maximum of 64 taxa in karst springs. A substantial proportion of the community was recorded only once in the dataset (n=63), representing half of all identified taxa. Some species formed well-developed populations in specific springs (*Humidophilla contenta* (Grunow) R.L.Lowe, Kociolek, Johansen, Van de Vijver, Lange-Bertalot and Kopalová in karst springs or *Tabellaria floculosa* (Roth) Kützing in crystalline-metamorphic springs), whereas other species appeared as a few frustules across samples (*Amphipleura pellucida* (Kützing) Kützing, *Gyrosigma nodiferum* (Grunow) Reimer, *Stephanocyclus meneghinianus* (Kützing) Kulikovskiy, Genkal and Kociolek).

A contrasting pattern was observed for *Achnanthidium minutissimum* (Kützing) Czarnecki, *Planothidium lanceolatum* (Brébisson ex Kützing) Lange-Bertalot and *Meridion circulare* (Greville) C.Agardh, which occurred in all ten springs, alongside other frequent species like *Cocconeis placentula* Ehrenberg and *Gomphonema parvulum* (Kützing) Kützing. Singleton taxa prevailed in crystalline springs (n=41), with a high number of taxa belonging to genera *Eunotia* and *Pinnularia*, of which 17 taxa were identified in samples from SR10. In karstic springs only 19 taxa were identified with a single occurrence throughout the analyzed samples (i.e. *Gomphonema truncatum* Ehrenberg, *Navicula menisculus Schumann*, *Nitzschia palea* (Kützing) W.Smith).

Community differentiation among springs

Species composition exhibited limited overlap among springs: Jaccard similarity values indicating substantial compositional turnover among springs (Supplementary Figure 1). The crystalline spring SR10 stood out as the most distinctive, sharing only 15% of its taxa with the other springs. The greatest resemblance between the two groups was detected between GR7 and BC9 (43%), both being rheocrene springs.

The relative abundance analysis highlighted *Achnanthidium minutissimum* as the most consistently present taxon, which predominated in six of the 10 springs. This species reached high proportions (up to 68%) and dominated assemblages in crystalline springs, except SR10. The latter exhibited a different dominance pattern with *Eunotia exigua* (Brébisson ex Kützing) Rabenhorst accounting for more than 40% of the samples. Another particular pattern was observed in karst spring GR6, species *Humidophilla contenta* prevailing in both epibryon and epilithon samples. A further taxon capable of attaining dominance and, in some cases, competing with *A. minutissimum* was *Amphora pediculus*, which reached up to 50% relative abundance in GR7 and occurred at lower percentages in several other karst springs. *Planothidium lanceolatum* was also widespread in the dataset and contributed substantial proportions in the crystalline springs (BC8 and BC9).

Diatom diversity (Shannon index) reached a maximum value of 2.04 in samples from karst springs (GR5), where taxon evenness was relatively high (mean 0.58) and dominance low (0.26). Notably, samples from GR6 exhibited the highest dominance index, reflecting the strong prevalence of *H. contenta* on the available substrate relative to other taxa. The diversity of crystalline-metamorphic springs reached a maximum of $H' = 1.93$, despite being represented by a single sample each. Shannon diversity varied only marginally among springs ($F_{10,8} = 2.60$, $p = 0.095$), indicating relatively consistent alpha diversity among springs.

A strong representation of diatom community patterns across the studied springs was provided by a two-dimensional NMDS ordination based on Bray–Curtis dissimilarity (stress = 0.045; Fig. 2). The ordination plot showed a clear tendency for samples to cluster according to geological substrate. Carbonate-dominated springs formed a relatively coherent group, while crystalline–metamorphic springs displayed broader dispersion along the main ordination axes. The high goodness-of-fit values (non-metric $R^2 = 0.998$; linear $R^2 = 0.994$) further confirm that the two-dimensional NMDS solution provides an excellent representation of the underlying dissimilarity structure.

Given the complexity of geological influences on diatom assemblages, measured environmental parameters were projected onto the ordination (Fig. 2). This analysis revealed a strong association between crenic diatom communities and water chemistry. Environmental fitting indicated significant associations between community composition and dissolved oxygen (r^2 = 0.57, p = 0.008), conductivity (r^2 = 0.48, p = 0.007), pH (r^2 = 0.47, p = 0.014), and canopy cover treated as an ordinal variable (r^2 = 0.59, p = 0.002). Whereas temperature showed a near-significant association (p = 0.056), while flow intensity treated as ordinal was not significant (p = 0.237). When fitted as a categorical factor, geological substrate was significantly associated with the ordination (r^2 = 0.195, p = 0.029), indicating compositional differentiation among geological groups.

Results indicate that geological substrate contributes to community differentiation, while hydrochemical and shading-related gradients further structure assemblage variation within and across geological settings.

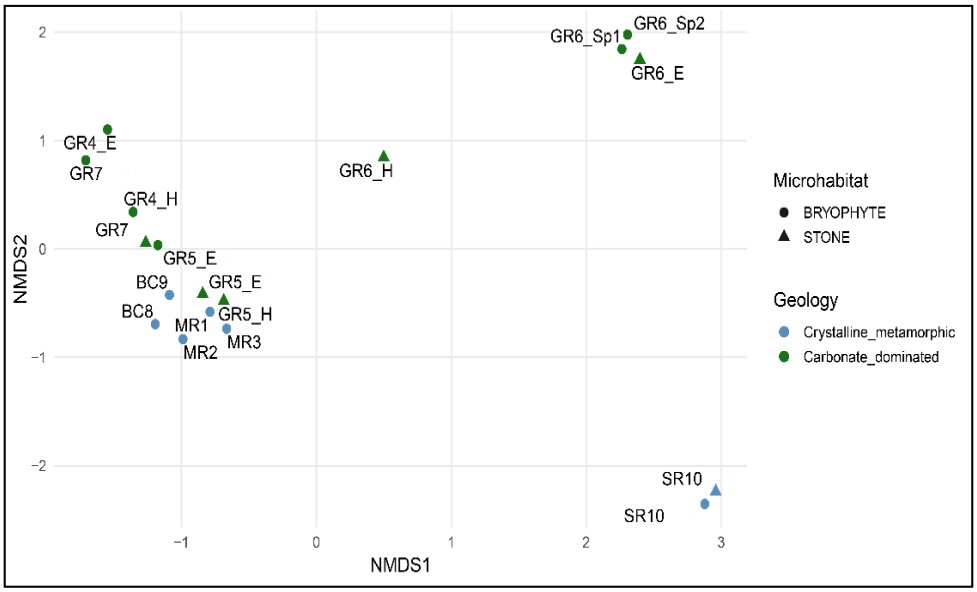


Figure 2. Non-metric multidimensional scaling (NMDS) ordination of diatom communities based on Bray–Curtis dissimilarities (stress = 0.045). Symbol shapes indicate the sampled microhabitat (circles – epibryon; triangles – epilithon) and color indicates geological substrate (green – carbonate-dominated; blue – crystalline-metamorphic). Other abbreviations are provided in Table 1.

Geological and environmental structuring of community composition

Diatom assemblages differed significantly between carbonate-dominated and crystalline–metamorphic springs. In the full dataset, PERMANOVA based on Bray–Curtis dissimilarities revealed a significant effect of geological setting on community composition (R^2 = 0.138, F = 2.56, p = 0.02, 999 permutations). Homogeneity of multivariate dispersion did not differ between geological groups (p = 0.422), indicating that the observed differences reflect genuine compositional shifts rather than unequal within-group variability. When tested as a single predictor, geology significantly explained community variation (RDA: F = 2.64, p = 0.03, 999 permutations; adjusted R^2 = 8.8%), accounting for 14.2% of the total variance. It is consistent with the NMDS clustering pattern and supports a detectable and consistent geological influence on community composition.

Furthermore, an exploratory redundancy analysis (RDA) was conducted using the full set of measured predictors and revealed that community composition was significantly structured by the combined set of geological and environmental predictors, explaining 61.2% of total variance (overall model: p = 0.002; adjusted R^2 = 0.34). After conducting marginal tests, two predictors remained significant: canopy cover (p = 0.026) and pH with a near-significant contribution (p = 0.059). Part of the geological effect detected in PERMANOVA was mediated through associated hydrochemical gradients, as indicated by the p-value of 0.141 obtained after accounting for the measured environmental variables. Moreover, variance inflation factors in the full model were below commonly accepted thresholds (all VIF < 5), but conductivity and temperature showed overlapping explanatory patterns with pH and canopy, suggesting partial redundancy among environmental predictors. Given the limited sample size relative to the number of predictors, and to avoid over-parameterization, a reduced model was constructed retaining variables which were ecologically interpretable and statistically supported.

Variation partitioning indicated that geological and environmental predictors together explained 23.8% of the variation in diatom community composition (adjusted R^2). Environmental variables accounted for the largest independent fraction (15.0%), whereas geology explained only a small independent component (1.6%), with an additional shared fraction (7.2%) reflecting the indirect influence of geology through hydrochemical gradients.

Although the independent geological fraction was small, the shared component with environmental variables indicates that geological effects on diatom assemblages are mediated through hydrochemical gradients related to bedrock composition.

The reduced RDA model (Fig. 3) including pH, dissolved oxygen, canopy cover, and geology remained significant (overall p = 0.002), explaining 52.5% of total variance (adjusted R^2 = 0.33). Marginal tests in the reduced model were consistent with the previous outcome, canopy cover (p = 0.020) and pH (p = 0.046) significantly structured assemblage composition, whereas dissolved oxygen and geology did not retain independent significance. Absence of multicollinearity was confirmed by low values of variance inflation factors.

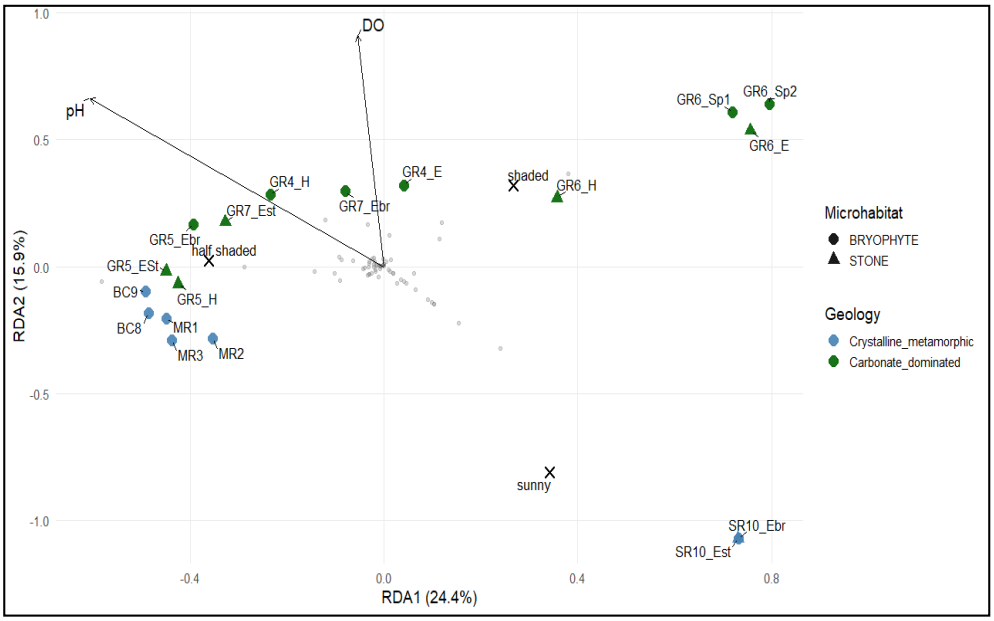


Figure 3. Redundancy analysis (RDA) ordination of diatom assemblages from karst and crystalline springs, based on the reduced model. The first two constrained axes were significant with explained variation for axis 1: 24.4% (p = 0.010) and axis 2: 15.9% (p = 0.013). Symbol shapes indicate the sampled microhabitat (circles – epibryon; triangles – epilithon) and color indicates geological substrate (green – carbonate-dominated; blue – crystalline-metamorphic), centroid position of canopy cover variable is indicated with x shape. Other abbreviations are provided in Table 1.

Together, these results further reinforce that while geological substrate exerts a detectable influence on community composition, much of its effect is mediated through associated environmental gradients, particularly light availability and hydrochemical conditions (pH), rather than as an independent categorical driver.

Ecological indicator coherence and functional interpretation of assemblage patterns

Across springs in both geological contexts, assemblages were dominated by low-profile taxa, which typically accounted for 65–100% of relative abundance. High-profile and motile guilds varied, but generally remained secondary components, while planktonic taxa were rare. One crystalline–metamorphic spring (SR10) stood out with a distinct configuration dominated by high-profile taxa (80%) and containing notable planktonic forms (15%). This pattern corresponded to its position in the ordination and to its ecological indicator profile, indicating environmental conditions distinct from the other sites. This is consistent with the slightly different crystalline siliceous metasedimentary bedrock and its distinct hydrochemical signature (very low conductivity and comparatively low oxygenation).

Van Dam ecological indicator profiles were consistent with the lithological contrast observed among springs. The pH spectrum coincides with the measured values (7.34-7.69); alkaline indicators representing 71.7% on average, compared with 28.2% neutrophilic indicating taxa. Oxygen preferences were broadly consistent with well-oxygenated conditions characteristic of karst springs with medium to high flow (high-O_2 classes: 73.7%), aligning with the measured hydrochemistry of above 9.4 mg/L of dissolved oxygen. The diatoms sensitive to moisture levels were dominated by taxa classified as "aquatic to aerophilic" (51%) in the carbonate-dominated group, with the highest contribution of aerophilic (21%) and occasionally aerophilic (24.4%) taxa, compared to the opposite group. These values reflect the frequent fluctuations that occur in karst springs, especially at spring GR6, which is known to experience a marked decrease in discharge over the course of the year, sometimes drying out completely for several days (Orăşeanu, 2016).

In contrast, crystalline sites displayed a different indicator species distribution. These patterns support geology as an important background factor influencing community composition. In the pH spectrum, communities shifted toward neutrophilic dominance (59.4%) with substantially lower alkaline indicators (25.4%) and a much larger acid-indicator component (maximum of 82.8%). Oxygen preferences remained similar to karst at the group level (78.0% indicating high-O_2 classes). Overall, the karst–crystalline separation is expressed most strongly through the presence of a high percentage of acidophilic taxa and a dominant composition of aquatic to aerophilic taxa (88%).

Indicator species analysis (IndVal.g) identified three taxa significantly associated with the geological groups. *Pinnularia subcapitata* was associated with crystalline–metamorphic springs (IndVal.g stat = 0.655, p = 0.046), whereas

Meridion circulare (stat = 0.905, p = 0.012), *Cocconeis placentula* (stat = 0.795, p = 0.048) and *Cocconeis euglypta* (stat = 0.798, p = 0.039) were associated with carbonate-dominated springs.

These indicator patterns support non-random ecological differentiation consistent with the geological contrast, and coherent with results obtained by environmental fitting analysis and RDA. Furthermore, these results support the hypothesis that geological substrate influences diatom assemblage structure and justify the focused investigation of karst springs conducted in subsequent studies.

Discussion

The species richness identified across the samples collected from the studied crenic sites was high, with a pronounced compositional heterogeneity among sites. Singleton taxa represented half of the total community; their numbers prevailing in crystalline-metamorphic springs. Community similarity between sites was generally low, and ordination analysis revealed a tendency for samples to group according to geological substrate. Multivariate analyses confirmed that diatom assemblage composition differed significantly between carbonate-dominated and crystalline–metamorphic springs. Geological substrate explained a moderate portion of the community variation, while measured environmental variables, particularly light availability and pH, showed significant associations with community structure. Variation partitioning showed that much of the variation attributed to geology was shared with environmental variables, indicating that geology structures diatom communities mainly through associated hydrochemical gradients. Functional analysis revealed low-profile taxa dominated in karstic springs, typical of stable benthic habitats, while high-profile taxa appeared mostly in crystalline-metamorphic springs. Ecological indicator values and indicator species analysis supported the differentiation of communities between geological contexts and highlighted the influence of local hydrochemical and hydrological conditions in both spring groups.

The patterns observed in the present study are consistent with those reported by Cantonati (1998) for mountain springs in the Southern Alps, where diatom assemblages were primarily structured by hydrochemical gradients closely linked to geology. Results show that carbonate springs exhibited higher pH, conductivity and alkalinity, whereas siliceous springs were characterized by lower mineralization and slightly more acidic conditions, resulting in distinct species assemblages. Similarly, our results suggest that geology influences crenic diatom communities mainly through its control on water chemistry rather than

acting as a direct ecological filter. In particular, the differentiation of spring groups along conductivity and pH gradients, together with the strong association between conductivity and geological substrate, supports the view that hydrochemical conditions mediate the relationship between geology and diatom community composition. Cantonati (1998) also emphasized that most taxa recorded in springs are widespread rheophilous species typical of cool, flowing headwaters, while strictly crenophilous species represent only a small fraction of the flora. This observation is consistent with our findings, where dominant taxa such as *Meridion circulare* and other common benthic species are characteristic of well-oxygenated, flowing microhabitats frequently associated with bryophytes in spring environments.

Similar findings from Alpine studies identified geology as an important factor contributing to benthic diatom community composition, mainly through its influence on ionic composition and buffering capacity (Aboal *et al*., 1998; Cantonati *et al*., 2006; Cantonati *et al*., 2012b). In crenic habitats, hydrochemical parameters such as conductivity and alkalinity have also been shown to shape diatom communities through the influence of geological substrate (Cantonati and Spitale, 2009). This pattern is further supported by Mediterranean karst springs, where diatom assemblages were strongly associated with mineral content and carbonate geology (Lai *et al*., 2016), and by Beracko *et al*. (2022), who found that water chemistry explained more variation in diatom assemblages than substrate alone. Our findings therefore align with broader European evidence, indicating that geology mainly acts by shaping environmental gradients rather than functioning as a direct ecological filter.

Comparable patterns have been documented in protected areas in the Alps, where hydrochemical contrasts and substratum types are more pronounced (Cantonati and Spitale, 2009). The Mediterranean karst systems similarly show distinct assemblages among springs even within the same geological region (Lai *et al*., 2016). Studies from other regions, including Iranian springs, report that conductivity and nutrient gradients generate clear compositional differentiation at the spring site (Ahmadi Musaabad *et al*., 2019).

Indicator species analysis revealed *Pinnularia subcapitata* as an indicating species of crystalline metamorphic lithological substrate of a crenic aquatic habitat. This species usually supports an interpretation of low buffering and acid influence habitats (Poulíčková *et al*., 2005), which reflects the chemical profile exhibited by the measured variables at the eucrenal area of the crystalline siliceous spring.

Meridion circulare is typically associated with cool, well-oxygenated, flowing microhabitats in spring ecosystems, where it frequently develops on stable substrates and bryophyte-associated niches (Wojtal and Sobczyk, 2012). The

species is commonly reported from circumneutral to slightly alkaline waters and moderately mineralized karst springs, conditions often linked to carbonate bedrock influence (Lai *et al*., 2019). In this context, the occurrence and relative abundance of *M. circulare* in our dataset are consistent with the hydrochemical conditions typical of carbonate-dominated springs, supporting the interpretation that geology influences diatom assemblages primarily through its effect on water chemistry and flow-related microhabitat conditions.

Cocconeis euglypta is an adnate, low-profile periphytic diatom, typically associated with stable substrates in well-oxygenated, flowing environments, where its growth form allows it to tolerate moderate turbulence. In spring ecosystems, it is often frequent or abundant across various substrates, including bryophytes, and has been linked to mineral-rich karst springs where hydrochemical gradients and discharge variability influence community structure (Cantonati, 1998). These indicating environmental characteristics are consistent with the habitats in which we identified this species in high abundances.

The frequent occurrence of taxa such as *Amphora pediculus* and *Cocconeis placentula* is consistent with previous studies on karst springs in the Apuseni Mountains, where these species were identified as dominant components of crenic diatom communities (Şuteu *et al*., 2025). *Amphora pediculus*, an alkaliphilous and relatively tolerant species, was identified in our study in moderately to highly mineralized springs and mesotrophic to eutrophic conditions. This taxon alongside the frequent *Cocconeis placentula*, are typically linked to stable lithic substrates in carbonate-influenced systems (Cantonati *et al*., 2012b).

The inclusion of both epilithic and bryophytic microhabitats likely contributed to the relatively high taxonomic richness observed across sites, as these substrates are known to support partially distinct diatom assemblages in spring ecosystems (Cantonati *et al*., 2012a). Site samples differentiation was detectable among springs, but this pattern was largely driven by a small number of ecologically distinct sites, while karst springs shared similar species composition.

The exceptionally low conductivity recorded at Someşul Rece (SR10 = 13 $\mu S\ cm^{-1}$) is consistent with the siliceous metasedimentary bedrock (phyllites and quartzites) of the Pts_Pz unit (a Paleozoic metamorphic lithological unit in the national geological map), which contributes only limited dissolved ions due to the slow weathering of silicate minerals and the absence of carbonate buffering. The diatom community was characterized by a distinct dominance structure driven by *Eunotia exigua*, a well-known acidophilous taxon typically associated with low-conductivity, weakly mineralized waters, reflecting the siliceous substrate and reduced buffering capacity characteristic of crystalline systems. Thus, despite belonging to the same geological group as the other crystalline

springs, the distinct community structure observed at SR10 likely reflects locally expressed hydrochemical conditions derived from the siliceous substrate rather than anthropogenic influence or episodic disturbance.

Another similar phenomenon appeared in samples from the karst spring GR7, which exhibited a different ecological signal, dominated by *Humidophilla contenta*. It is a species indicative of intermittently wetted habitats and microhabitats experiencing periodic desiccation or fluctuating moisture conditions (Cantonati *et al*., 2012b; Lowe *et al*., 2014). Its prevalence in both epibryon and epilithon samples suggests localized micro-environmental instability or periodic moisture fluctuations within the spring habitat, consistent with the hydrological instable conditions identified throughout the year at this spring site (Orășeanu, 2016).

Van Dam ecological classification further clarified community structure. Most assemblages were dominated by taxa classified as strictly aquatic or aquatic-to-aerophilic, indicating stable flowing conditions typical of perennial crenic environments. However, the presence of occasionally aerophilic and aerophilic taxa, suggests localized moisture variability and hydrological dynamics. Such taxa have also been associated with fluctuating discharge in Bavarian springs (Cantonati *et al*., 2009) and with intermittent or seasonally variable karst systems (Orășeanu, 2016).

Research on spring diatoms in the Carpathian region has developed more gradually compared with the extensive investigations conducted in the Alps. Existing studies have contributed valuable insights into the diversity and ecology of diatom assemblages in mountain springs and spring fens (Pouličková *et al*., 2005; Micle *et al*. 2018, Şuteu *et al*., 2025).

Although further research is needed to fully understand the complexity of the Carpathian springs ecosystems, this study provides an initial step toward disentangling the relationships among geology, hydrochemistry, and diatom community composition by integrating environmental data, microhabitat structure, ordination analyses, and ecological indicator approaches. We further contribute to reducing the existing knowledge gap by presenting, for the first time, diatom-based results from three previously unstudied springs.

Conclusions

Diatom communities in the investigated springs were primarily structured by hydrochemical gradients associated with geological substrate, particularly those reflected in conductivity and pH. Multivariate analyses indicated clear differences in community composition between carbonate-dominated and crystalline–metamorphic springs. Although geological substrate contributed to

the observed variation in diatom community structure, its influence was expressed mainly through environmental gradients linked to water chemistry. Among the measured variables, pH and canopy cover emerged as important factors associated with differences in diatom community composition, highlighting the combined influence of hydrochemical conditions and local shading on crenic diatom flora.

The ecological characteristics of the recorded taxa and guild composition indicate stable conditions typical of freshwater spring environments, with ecologically meaningful signals of moisture variability and localized nutrient influence. In a broader European context, these patterns are consistent with findings from spring systems in the Alps and Mediterranean mountain region, emphasizing the role of hydrochemistry and geology in shaping diatom communities. At the same time, the results highlight the need for further integrative research on spring ecosystems of the Carpathian Mountains in order to better understand their ecological dynamics and potential responses to increasing environmental pressures.

Acknowledgements. The Apuseni Nature Park Administration is acknowledged for granting access permits to springs located within protected areas. We also thank the research team of the Advanced Hydrobiology and Biomonitoring Laboratory from Faculty of Biology and Geology, Babeș-Bolyai University, for providing the equipment necessary for this study.

Funding: not applicable.

Conflicts of Interest: The authors declare no competing interests.

Ethics declarations: not applicable.

References

Aboal, M., Puig, M.A. & Prefasi, M. (1998). Diatom assemblages in springs in Castellón province, eastern Spain. *Arch. Hydrobiol. Suppl. 125*, 79–95.

Ahmadi Musaabad, L., Panahy Mirzahasanlou, J., Mahmoodlu, M.G. & Bahlakeh, A. (2019). Diatom flora in three springs of Golestan Province. *J. Phycol. Res. 3*(2), 432–442.

Angeli, N., Cantonati, M., Spitale, D., & Lange-Bertalot, H. (2010). A comparison between diatom assemblages in two groups of carbonate, low-altitude springs with different levels of anthropogenic disturbances. *Fottea 10(1),* 115-128. https://doi.org/10.5507/FOT.2010.006

Baker, L.-A., Biron, D.G., Millan, F., Voldoire, O., Breton, V., Allain, E., Wetzel, C.E., Ector, L. & Beauger, A. (2022). The substrate, a key factor or not, to explain the species diversity of diatom communities in mineral springs. *Bot. Lett.* https://doi.org/10.1080/23818107.2022.2028186

Beauger, A., Voldoire, O., Allain, E., Gosseaume, P., Blavignac, C., Baker, L.-A. & Wetzel, C.E. (2023). Biodiversity and environmental factors structuring diatom assemblages of mineral saline springs in the French Massif Central. *Diversity 15*, 283. https://doi.org/10.3390/d15020283

Bellinger, E.G. & Sigee, D.C. (2010). Freshwater algae: Identification and use as bioindicators. *Wiley-Blackwell*, Chichester, pp. 284.

Beracko, P., Sochuliaková, L., Cíbik, J. & Lánczos, T. (2022). Drivers predicting biodiversity and community assembly of benthic diatoms in a karst spring environment. *Diatom Res. 37*, 181–197. https://doi.org/10.1080/0269249X.2022.2100486

Blanco, S., Viso, R., Borrego-Ramos, M., López-Flores, R., Mota-Echeandía, D., Tierra, M. & Castañeda, C. (2025). The ecology of benthic diatom assemblages in saline wetlands of the Ebro Basin, NE Spain. *Microb. Ecol. 88*, 16. https://doi.org/10.1007/s00248-025-02514-3

Cantonati, M. (1998). Diatom communities of springs in the Southern Alps. *Diatom Res. 13(2),* 201-220. https://doi.org/10.1080/0269249X.1998.9705449

Cantonati, M., & Pipp, E. (2000). Longitudinal and seasonal differentiation of epilithic diatom communities in the uppermost sections of two mountain spring-fed streams. Verh. - *Int. Ver. Theor. Angew. Limnol.* **27** (3), 1591–1595. https://doi.org/10.1080/03680770.1998.11901507

Cantonati, M., Gerecke, R., & Bertuzzi, E. (2006). Springs of the Alps–sensitive ecosystems to environmental change: from biodiversity assessments to long-term studies. *Hydrobiologia 562(1)*, 59-96. https://doi.org/10.1007/s10750-005-1806-9

Cantonati M., Rott E., Pfister P. & Bertuzzi E. (2007). Benthic algae in springs: biodiversity and sampling methods. In: Cantonati M., Bertuzzi E. & Spitale D. (eds), The spring habitat: biota and sampling methods. Ed. *Museo Tridentino di Scienze Naturali*, Trento: 77-112 (Monografi e del Museo Tridentino di Scienze Naturali, 4).

Cantonati, M. & Spitale, D. (2009). The role of environmental variables in structuring epiphytic and epilithic diatom assemblages in springs and streams of the Dolomiti Bellunesi National Park (south-eastern Alps). *Fundam. Appl. Limnol. 174*, 117–133. https://doi.org/10.1127/1863-9135/2009/0174-0117

Cantonati, M., Füreder, L., Gerecke, R., Jüttner, I. & Cox, E.J. (2012a). Crenic habitats, hotspots for freshwater biodiversity conservation: Toward an understanding of their ecology. *Freshw. Sci. 31*, 463–480. https://doi.org/10.1899/11-111.1

Cantonati, M., Angeli, N., Bertuzzi, E., Spitale, D., & Lange-Bertalot, H. (2012b). Diatoms in springs of the Alps: spring types, environmental determinants, and substratum. *Freshw. Sci. 31(2)*, 499-524. https://doi.org/10.1899/11-065.1

Cantonati, M., Bilous, O., Spitale, D., Angeli, N., Segadelli, S., Bernabè, D., Lichtenwöhrer, K., Gerecke, R. and Saber, A.A. (2022). Diatoms from the spring ecosystems selected for the long-term monitoring of climate-change effects in the Berchtesgaden National Park (Germany). *Water*, *14*(3), 381. https://doi.org/10.3390/w14030381

De Cáceres, M., & Legendre, P. (2009). Associations between species and groups of sites: Indices and statistical inference. *Ecology*, 90(12), 3566–3574. https://doi.org/10.1890/08-1823.1

De Cáceres, M. (2023). *indicspecies:* Relationship between species and groups of sites (R package version 1.7.14). https://cran.r-project.org/package=indicspecies

Fernández-Martínez, M., Barquín, J., Bonada, N., Cantonati, M., Churro, C., Corbera, J. & Romero, E. (2024). Mediterranean springs: Keystone ecosystems and biodiversity refugia threatened by global change. *Glob. Change Biol. 30*, e16997. https://doi.org/10.1111/gcb.16997

Ford, D.C. & Williams, P.W. (2007). Karst hydrogeology and geomorphology. *John Wiley & Sons*, Chichester, pp. 562.

Fránková, M., Bojková, J., Poulíčková, A., & Hájek, M. (2009). The structure and species richness of the diatom assemblages of the Western Carpathian Spring fens along the gradient of mineral richness, *Fottea 9*(*2*), 355-368. https://doi.org/10.5507/fot.2009.035

Geological Institute of Romania (revised 2018). Geology, INSPIRE Annex II, scale 1:200,000 [WMS service]. Geological Institute of Romania, Bucharest [Accessed 01 March 2026] https://inspire-geoportal.ec.europa.eu

Hájková, P., Bojková, J., Franková, M., Opravilová, V., Hájek, M., Kintrová, K. & Horsák, M. (2011). Disentangling the effects of water chemistry and substratum structure on moss-dwelling unicellular and multicellular micro-organisms in spring fens. *J. Limnol. 70*, 54–60. https://doi.org/10.4081/jlimnol.2011.s1.54

Hammer, Ø., Harper, D.A.T., & Ryan, P.D. (2001). PAST: Paleontological Statistics software package for education and data analysis. *Palaeontol*. *Electron*. *4*(*1*): 9. https://doi.org/10.1016/j.ecohyd.2020.07.007

Kelly, M.G., Cazaubon, A., Coring, E., Dell'Uomo, A., Ector, L., Goldsmith, B., Guasch, H., Hürlimann, J., Jarlman, A., Kawecka, B., Kwandrans, J. (1998). Recommendations for the routine sampling of diatoms for water quality assessments in Europe. *J. Appl. Phycol.*, *10*(*2*), 215-224. https://doi.org/10.1023/A:1008033201227

Krammer, K., Lange-Bertalot, H. (1986). Bacillariophyceae, 1. Teil: Naviculaceae, In: Ettl H, Gerloff, J., Heynig, H., Mollenhauer, D. (eds), Süßwasserflora von Mitteleuropa 2/1. Ed. *Gustav Fischer Verlag*, Stuttgart & New York, pp. 876.

Krammer, K., Lange-Bertalot, H. (1988). Bacillariophyceae, 2. Teil: Bacillariaceae, Epithemiaceae, Surirellaceae, In: Ettl, H., Gerloff, J., Heynig, H., Mollenhauer, D. (eds), Süßwasserflora von Mitteleuropa 2/2. Ed. *Gustav Fischer Verlag*, Stuttgart & New York, pp. 596.

Krammer, K., Lange-Bertalot, H. (1991a). Bacillariophyceae, 3. Teil: Centrales, Fragilariaceae, Eunotiaceae, In: In: Ettl, H., Gerloff, J., Heynig, H., Mollenhauer, D. (eds), Süßwasserflora von Mitteleuropa 2/3. Ed. *Gustav Fischer Verlag*, Stuttgart, Jena, pp. 598.

Krammer, K., Lange-Bertalot, H. (1991b). Bacillariophyceae, 4. Teil: Achnanthaceae. Kritische Ergänzungen zu Navicula (Lineolatae) und Gomphonema, In: Ettl, H., Gärtner, G., Gerloff, J., Heynig, H., Mollenhauer, D. (eds), Süßwasserflora von Mitteleuropa 2/4. Ed. *Gustav Fischer Verlag*, Stuttgart & New York, pp. 437.

Lai, G. G., Padedda, B. M., Wetzel, C. E., Lugliè, A., Sechi, N., & Ector, L. (2016). Epilithic diatom assemblages and environmental quality of the Su Gologone karst spring (centraleastern Sardinia, Italy). *Acta Bot. Croat. 75*(*1*), 129-143.

Lai, G.G., Burato, S., Padedda, B.M., Zorza, R., Pizzul, E., Delgado, C. & Cantonati, M. (2019). Diatom biodiversity in karst springs of Mediterranean geographic areas with contrasting characteristics: Islands vs mainland. *Water 11*, 2602. https://doi.org/10.3390/w11122602

Lai, G.G., Padedda, B.M., Ector, L., Wetzel, C.E., Lugliè, A. & Cantonati, M. (2020). Mediterranean karst springs: Diatom biodiversity hotspots under the pressure of hydrological fluctuation and nutrient enrichment. *Plant Biosyst. 154*, 673–684. https://doi.org/10.1080/11263504.2019.1674402

Lecointe, C., Coste, M. & Prygiel, J. (1993). "Omnidia": Software for taxonomy, calculation of diatom indices and inventories management. *Hydrobiologia 269*, 509–513. https://doi.org/10.1007/BF00028048

Letáková, M., Fránková, M. & Poulíčková, A. (2018). Ecology and applications of freshwater epiphytic diatoms. *Cryptogamie Algol. 39*, 3-22. https://doi.org/10.7872/crya/v39.iss1.2018.3

Lowe, R.L., Kociolek, J.P., Johansen, J.R., Van de Vijver, B., Lange-Bertalot, H. & Kopalová, K. (2014). Humidophila gen. nov., a new genus for a group of diatoms (Bacillariophyta) formerly within the genus Diadesmis: species from Hawai'i, including one new species. *Diatom Res.* 29, 351-360. https://doi.org/10.1080/0269249X.2014.889039

Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J. & Bowler, C. (2016). Insights into global diatom distribution and diversity in the world's ocean. *Proc. Natl. Acad. Sci. U.S.A. 113*, E1516–E1525. https://doi.org/10.1073/pnas.1509523113

Micle, M., Şuteu, A. M., Momeu, L., Coman, P., Battes, K. P., & Cîmpean, M. (2018). Seasonal dynamics of biota from several springs located in the Apuseni Mountains, Romania. *Sci. Stud. Res. Biol. 27*(2).

Natural Earth (2024). Free vector and raster map data [Accessed 01 March 2026] https://www.naturalearthdata.com/

Nicolosi Gelis, M. M., Sathicq, M. B., Jupke, J., & Cochero, J. (2022). DiaThor: R package for computing diatom metrics and biotic indices. *Ecological Modelling, 465*, 109859. https://doi.org/10.1016/j.ecolmodel.2021.109859

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2024). *vegan: Community ecology package* (R package version 2.6-6). https://cran.r-project.org/package=vegan

Orăşeanu, I. (2016). Hidrogeologia carstului din Muntii Apuseni. 1st Ed. *Belvedere*, Oradea, pp. 289. ISBN 978-606-94188-0-2

Orăşeanu, I. (2020). Hidrogeologia carstului din Munţii Apuseni. 2nd Ed. *Belvedere*, Oradea, pp. 349. ISBN 978-606-94188-9-5.

Pebesma, E. (2018). Simple features for R: Standardized support for spatial vector data. *The R Journal, 10*(1), 439–446. https://doi.org/10.32614/RJ-2018-009

Pouličková, A., Hájková, P., Křenková, P. & Hájek, M. (2004). Distribution of diatoms and bryophytes on linear transects through spring fens. *Nova Hedwigia 78*, 411–424. https://doi.org/10.1127/0029-5035/2004/0078-0411

Pouličková, A., Hájek, M. & Rybníček, K. (eds.) (2005). Ecology and palaeoecology of spring fens of the West Carpathians. *Palacký University Press*, Olomouc, pp. 210.

Rimet, F., & Bouchez, A. (2012). Life-forms, cell-sizes and ecological guilds of diatoms in European rivers. *Knowl manag Aquat Ec*, (*406*), 01. https://doi.org/10.1051/kmae/2012018

Soininen, J. (2007). Environmental and spatial control of freshwater diatoms—A review. Diatom Res. 22, 473-490. https://doi.org/10.1080/0269249X.2007.9705724

Şuteu, A.-M., Momeu, L., & Puşcaş, M. (2025). The diatom communities from Apuseni Mountains: A first approach on crenic diatom flora. *Stud. Univ. Babes-Bolyai Biol., 70*(2), 163–195. https://doi.org/10.24193/subbbiol.2025.2.10

Taxböck, L., Linder, H.P. & Cantonati, M. (2017). To what extent are Swiss springs refugial habitats for sensitive and endangered diatom taxa? *Water 9*, 967. https://doi.org/10.3390/w9120967

Van Dam, H., Mertens, A., & Sinkeldam, J. (1994). A coded checklist and ecological indicator values of freshwater diatoms from the Netherlands. *Neth. J. Aquat. Ecol*. *28*(1), 117-133. https://doi.org/10.1007/BF02334251

Vanormelingen, P., Verleyen, E. & Vyverman, W. (2008). The diversity and distribution of diatoms: From cosmopolitanism to narrow endemism. *Biodivers. Conserv. 17*, 393–405. https://doi.org/10.1007/s10531-007-9253-4

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. Springer, Cham, pp. 260. https://doi.org/10.1007/978-3-319-24277-4

Wojtal, A.Z. & Sobczyk, Ł. (2006). Composition and structure of epilithic diatom assemblages on stones of different size in a small calcareous stream (S Poland). *Algol. Stud*. 119, 105-124. https://doi.org/10.1127/1864-1318/2006/0119-0105

Wojtal, A. Z., & Sobczyk, Ł. (2012). The influence of substrates and physicochemical factors on the composition of diatom assemblages in karst springs and their applicability in water-quality assessment. *Hydrobiologia 695*(1), 97-108. https://doi.org/10.1016/j.ecolind.2011.07.002

Supplementary Table 1. Taxonomic list of diatoms recorded in the studied springs, with presence/absence data across epibryon and epilithon samples from each sampling site. The symbol "–" indicates absence, whereas "1" indicates presence

Taxa	**Author**	**MR1**	**MR2**	**MR3**	**GR4**	**GR5**	**GR6**	**GR7**	**BC8**	**BC9**	**SR10**	**GR5**	**GR6**	**GR7**	**SR10**
		Epibryon										**Epilithon**			
Achnanthidium minutissimum	(Kützing) Czarnecki	1	1	1	1	1	1	1	1	1	1	1	1	1	-
Amphipleura pellucida	(Kützing) Kützing	-	-	-	-	1	-	-	-	-	-	1	-	-	-
Amphora libyca	Ehrenberg	1	-	1	1	1	-	-	-	-	-	1	-	-	-
Amphora ovalis	(Kützing) Kützing	-	-	-	1	1	-	-	-	1	-	1	-	-	-
Amphora pediculus	(Kützing) Grunow	1	1	-	1	1	-	1	1	1	-	1	-	1	-
Aulacoseira alpigena	(Grunow) Krammer	-	-	-	-	-	-	-	-	-	1	-	-	-	1
Aulacoseira nivalis	(W.Smith) J.English & Potapova	-	-	-	-	-	-	-	-	-	1	-	-	-	-
Brachysira brebissonii	R.Ross	-	-	-	-	1	-	-	-	-	1	-	-	-	1
Brachysira vitrea	(Grunow) R.Ross	1	-	-	-	-	-	-	-	-	-	-	-	-	-
Caloneis silicula	(Ehrenberg) Cleve	-	-	-	-	1	-	-	-	-	-	-	-	-	-
Caloneis tenuis	(W.Gregory) Krammer	-	-	-	1	-	-	-	1	-	-	-	-	-	-
Cavinula pseudoscutiformis	(Hustedt) D.G.Mann & Stickle	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Cocconeis lineata	Ehrenberg	-	-	-	1	1	-	-	-	-	-	1	-	1	-

Taxa	Author	MR1	MR2	MR3	GR4	GR5	GR6	GR7	BC8	BC9	SR10	GR5	GR6	GR7	SR10
		Epibryon										Epilithon			
Cocconeis placentula	Ehrenberg	1	-	1	1	1	1	1	1	1	1	1	1	1	-
Cocconeis euglypta	Ehrenberg	-	-	-	1	1	1	1	-	-	-	-	-	1	-
Cymbella aspera	(Ehrenberg) Cleve	1	-	-	1	1	-	-	-	-	-	-	-	-	-
Cymbella helvetica	Kützing	-	-	-	-	1	-	-	-	-	-	1	-	-	-
Cymbella lanceolata	C.Agardh	-	-	-	1	-	-	-	-	-	-	-	-	-	-
Cymbopleura naviculiformis	(Auerswald ex Heiberg) Krammer	1	-	-	-	1	-	-	-	-	-	-	-	-	1
Denticula tenuis	Kützing	-	-	-	-	1	-	-	-	-	1	1	-	-	-
Diatoma ehrenbergii	Kützing	-	-	-	-	1	-	-	-	-	1	1	-	-	-
Diatoma moniliformis	(Kützing) D.M.Williams	-	-	-	1	-	-	-	-	-	-	1	-	-	-
Diatoma vulgaris	Bory	-	-	-	-	1	1	-	-	-	-	1	1	-	-
Didymosphenia geminata	Mart.Schmidt	-	-	-	-	1	-	-	-	-	-	1	-	-	-
Diploneis elliptica	(Kützing) Cleve	-	-	-	-	-	-	-	1	-	-	-	-	-	-
Diploneis oblongella	(Nägeli ex Kützing) A.Cleve	-	1	1	-	1	-	-	1	-	-	1	-	-	-
Diploneis ovalis	(Hilse) Cleve	-	1	-	-	-	-	-	-	-	-	-	-	-	-
Encyonema gracile	Rabenhorst	-	-	-	-	1	-	-	-	-	1	1	-	-	1

Taxa	Author	MR1	MR2	MR3	GR4	GR5	GR6	GR7	BC8	BC9	SR10	GR5	GR6	GR7	SR10
		Epibryon										Epilithon			
Encyonema minutum	(Hilse) D.G.Mann	1	-	-	1	1	1	-	1	-	1	1	-	1	1
Encyonema silesiacum	(Bleisch) D.G.Mann	-	-	1	-	1	1	-	-	-	-	1	-	-	-
Eunotia bilunaris	(Ehrenberg) Schaarschmidt	1	1	-	-	-	-	-	1	-	1	-	-	-	1
Eunotia exigua	(Brébisson ex Kützing) Rabenhorst	-	1	-	-	-	-	-	1	1	1	1	1	-	1
Eunotia fallax	A.Cleve	-	-	-	-	-	-	-	1	-	1	-	-	-	1
Eunotia glacialis	F.Meister	-	-	-	-	-	-	-	-	-	1	-	-	-	1
Eunotia implicata	Nörpel, Lange-Bertalot & Alles	-	-	-	-	-	-	-	-	-	1	-	-	-	1
Eunotia mucophila	(Lange-Bertalot, Nörpel-Schempp & Alles) Lange-Bertalot	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Eunotia paludosa	Grunow	1	1	-	-	-	-	-	1	-	1	-	-	-	1
Eunotia praerupta	Ehrenberg	-	-	-	-	-	-	-	-	-	1	-	-	-	1
Eunotia tenella	(Grunow) Hustedt	-	-	-	-	-	-	-	-	-	1	-	-	-	1
Fragilaria capucina	Desmazières	-	1	-	-	1	-	-	-	-	-	1	-	-	-
Fragilaria tenera	(W.Smith) Lange-Bertalot	-	-	-	-	-	-	-	-	-	1	-	-	-	-
Fragilariforma virescens	(Ralfs) D.M.Williams & Round	-	-	-	-	-	-	-	1	-	1	1	-	-	1
Frustulia saxonica	Rabenhorst	-	-	-	-	1	-	-	-	-	1	-	-	-	1

Taxa	Author	MR1	MR2	MR3	GR4	GR5	GR6	GR7	BC8	BC9	SR10	GR5	GR6	GR7	SR10
		Epibryon										Epilithon			
Frustulia vulgaris	(Thwaites) De Toni	1	-	-	-	1	-	-	1	-	1	-	-	-	-
Gomphonella olivacea	(Hornemann) Rabenhorst	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Gomphonema acuminatum	Ehrenberg	1	-	-	-	-	-	-	-	-	-	-	-	-	-
Gomphonema angustatum	(Kützing) Rabenhorst	-	-	-	-	-	-	1	1	1	-	-	-	1	-
Gomphonema angustum	C.Agardh	-	-	-	-	1	-	1	1	-	-	1	-	-	-
Gomphonema gracile	Ehrenberg	1	-	-	-	-	-	-	-	-	-	1	-	-	-
Gomphonema parvulum	(Kützing) Kützing	1	-	1	-	-	1	-	1	1	1	1	1	1	1
Gomphonema subtile	Ehrenberg	1	-	-	-	-	-	-	-	-	-	-	-	-	-
Gomphonema truncatum	Ehrenberg	1	-	1	-	1	-	-	-	-	-	1	-	-	-
Gyrosigma nodiferum	(Grunow) Reimer	-										1			
Hannaea arcus	(Ehrenberg) R.M.Patrick	-	-	-	-	1	1	-	-	-	1	1	1	-	-
Hantzschia amphyoxis	(Ehrenberg) Grunow	1	-	-	-	1	-	-	1	1	-	-	-	1	-

Taxa	Author	MR1	MR2	MR3	GR4	GR5	GR6	GR7	BC8	BC9	SR10	GR5	GR6	GR7	SR10
		Epibryon										Epilithon			
Humidophila contenta	(Grunow) R.L.Lowe, Kociolek, Johansen, Van de Vijver, Lange-Bertalot & Kopalová	-	-	-	-	-	1	-	-	-	-	-	1	-	-
Humidophila perpusilla	(Grunow) R.L.Lowe, Kociolek, J.R.Johansen, Van de Vijver, Lange-Bertalot & Kopalová	-	1	1	-	-	-	-	1	1	-	1	-	1	-
Iconella linearis	(W.Smith) Ruck & Nakov	-	-	-	-	1	-	-	-	-	-	-	-	-	-
Luticola mutica	(Kützing) D.G.Mann	1	-	-	-	1	1	-	-	-	-	1	1	1	-
Melosira varians	C.Agardh	1	-	-	-	-	1	-	-	-	-	-	-	-	-
Meridion circulare	(Greville) C.Agardh	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Navicula capitata	F.E.Fritsch & F.Rich	1	-	-	-	-	-	-	-	-	-	-	-	-	-
Navicula capitatoradiata	H.Germain ex Gasse	-	-	-	-	-	1	-	-	-	-	-	-	-	-
Navicula cari	Ehrenberg	1	-	-	-	-	-	-	-	-	-	-	-	-	-
Navicula cincta	Pantocsek	-	-	-	1	1	1	-	-	-	-	1	-	-	-
Navicula cryptotenella	Lange-Bertalot	1	1	-	1	1	1	-	-	-	-	1	1	1	-
Navicula gregaria	Donkin	1	1	-	-	-	-	-	-	-	-	-	-	-	-
Navicula menisculus	Schumann	-	-	-	-	1	-	-	-	-	-	1	-	-	-
Navicula phyllepta	Kützing	-	-	-	-	1	-	-	1	-	1	1	-	-	1

Taxa	Author	MR1	MR2	MR3	GR4	GR5	GR6	GR7	BC8	BC9	SR10	GR5	GR6	GR7	SR10
		Epibryon										Epilithon			
Navicula radiosa	Kützing	1	1	-	1	1	-	-	-	-	-	-	-	-	-
Navicula recens	(Lange-Bertalot) Lange-Bertalot	-	-	-	-	1	-	-	-	-	-	1	-	-	-
Navicula rhyncocephala	Kützing	1	-	-	-	-	-	-	-	-	-	-	-	-	-
Navicula rostellata	Kützing	-	-	-	1	-	-	-	-	-	-	-	-	-	-
Navicula tripunctata	(O.F.Müller) Bory	-	-	-	1	1	1	1	-	1	-	1	1	1	-
Navicula trivialis	Lange-Bertalot	-	-	-	-	1	-	-	-	-	-	1	-	-	-
Navicula viridula	(Kützing) Ehrenberg	-	-	-	-	1	-	-	-	-	-	-	-	-	-
Neidium alpinum	Hustedt	-	1	-	-	-	-	-	-	-	-	-	-	-	-
Neidium ampliatum	(Ehrenberg) Krammer	-	-	-	-	-	-	-	1	-	1	-	-	-	1
Neidium bisulcatum	(Lagerstedt) Cleve	-	-	-	-	-	-	-	-	-	1	-	-	-	1
Neidium dubium	(Ehrenberg) Cleve	-	1	-	-	-	-	-	-	-	1	-	-	-	-
Nitzschia dissipata	(Kützing) Rabenhorst	1	-	-	-	1	1	-	-	-	-	1	-	-	-
Nitzschia intermedia	Hantzsch ex Cleve & Grunow	1	-	-	-	-	-	-	-	-	-	-	-	-	-
Nitzschia microcephala	Grunow	-	-	-	-	1	-	-	-	-	-	-	-	-	-
Nitzschia palea	(Kützing) W.Smith	1	1	1	-	1	-	-	1	-	-	1	-	-	-
Nitzschia paleacea	(Grunow) Grunow	1	-	-	-	1	-	-	1	-	-	-	-	-	-
Nitzschia perminuta	Grunow	-	-	-	-	-	-	-	1	1	-	-	-	-	-

Taxa	Author	MR1	MR2	MR3	GR4	GR5	GR6	GR7	BC8	BC9	SR10	GR5	GR6	GR7	SR10
		Epibryon										Epilithon			
Odontidium hyemale	(Roth) Kützing	1	-	-	-	-	-	-	-	-	-	-	-	-	-
Odontidium mesodon	(Ehrenberg) Kützing	1	1	-	1	1	-	1	1	1	-	1	-	-	-
Pantocsekiella ocellata	(Pantocsek) K.T.Kiss & Ács	-	-	-	-	-	1	-	-	-	-	-	-	-	-
Pinnularia abaujensis var. linearis	(Hustedt) R.M.Patrick	1	1	-	-	-	-	-	-	-	-	-	-	-	-
Pinnularia biceps	W.Gregory	-	-	-	-	-	-	-	-	-	1	-	-	-	-
Pinnularia borealis	Ehrenberg	-	1	-	-	-	-	-	-	-	1	-	-	-	1
Pinnularia divergens	W.Smith	-	1	-	-	-	-	-	-	-	-	-	-	-	-
Pinnularia divergentissima	(Grunow) Cleve	-	-	-	-	-	-	-	1	-	-	-	-	-	-
Pinnularia interrupta	W.Smith	-	1	-	-	-	-	-	-	-	-	-	-	-	-
Pinnularia isselana	Krammer	-	-	-	-	-	-	-	-	1	-	-	-	-	-
Pinnularia legumen	Ehrenberg	1	-	-	-	-	-	-	-	-	-	-	-	-	-
Pinnularia media	(Krammer) Kulikovskiy, Lange-Bertalot & Metzeltin	-	-	-	-	-	-	-	-	-	-	-	-	-	1

Taxa	Author	MR1	MR2	MR3	GR4	GR5	GR6	GR7	BC8	BC9	SR10	GR5	GR6	GR7	SR10
		Epibryon										Epilithon			
Pinnularia microstauron	(Ehrenberg) Cleve	-	-	-	-	1	-	-	-	-	1	-	-	-	1
Pinnularia nodosa	(Ehrenberg) W.Smith	-	1	-	-	-	-	-	-	-	-	-	-	-	-
Pinnularia obscura	Krasske	-	1	-	-	-	-	-	-	-	-	-	-	-	-
Pinnularia rupestris	Hantzsch	-	1	-	-	-	-	-	-	-	-	-	-	-	-
Pinnularia subcapitata	W.Gregory	-	1	-	-	-	-	-	1	-	1	-	-	-	1
Pinnularia sudetica	Hilse	-	-	-	-	-	-	-	-	1	-	-	-	-	-
Pinnularia viridis	(Nitzsch) Ehrenberg	-	1	-	-	-	-	-	1	-	1	-	-	-	1
Placoneis elginensis	(W.Gregory) E.J.Cox	-	1	1	-	-	-	-	-	-	-	-	-	-	-
Planothidium boyei	(Østrup) Van de Vijver, Coomans, Pottiez & C.E.Wetzel	-	-	-	1	-	-	-	1	1	-	-	-	-	-
Planothidium lanceolatum	(Brébisson ex Kützing) Lange-Bertalot	1	1	1	1	1	1	-	1	1	1	1	1	1	-
Rhoicosphaenia abbreviata	(C.Agardh) Lange-Bertalot	-	-	1	1	1	-	-	-	-	-	1	-	-	-
Rhopalodia gibba	(Ehrenberg) O.Müller	1	1	-	-	-	-	-	-	-	1	-	-	-	-
Sellaphora pupula	(Kützing) Mereschkovsky	-	-	-	-	-	-	-	-	-	1	1	-	-	-
Stauroneis anceps	Ehrenberg	-	-	-	-	1	-	-	-	-	1	-	-	-	1
Stauroneis phoenicenteron	(Nitzsch) Ehrenberg	1	-	-	-	-	-	-	-	-	1	-	-	-	-
Stauroneis smithii	Grunow	1	-	-	-	-	-	-	-	-	-	-	-	-	-

Taxa	Author	MR1	MR2	MR3	GR4	GR5	GR6	GR7	BC8	BC9	SR10	GR5	GR6	GR7	SR10
		Epibryon										Epilithon			
Staurosira construens	Ehrenberg	-	-	-	1	-	-	-	-	-	-	-	-	1	-
Staurosirella leptostauron	(Ehrenberg) D.M.Williams & Round	1	-	-	-	-	-	-	-	-	-	-	-	-	-
Staurosirella pinnata	(Ehrenberg) D.M.Williams & Round	-	-	-	-	-	-	-	-	-	1	-	-	-	1
Stephanocyclus meneghinianus	(Kützing) Kulikovskiy, Genkal & Kociolek	-	-	-	-	-	-	-	-	-	1	-	-	-	1
Surirella angusta	Kützing	1	-	-	-	1	1	-	1	-	-	1	-	-	-
Surirella elegans	Ehrenberg	1	-	-	-	-	-	-	-	-	-	-	-	-	-
Surirella librile	(Ehrenberg) Ehrenberg	-	-	-	-	1	-	-	-	-	-	-	-	-	-
Surirella microlibrile	Van de Vijver, Pottiez & Jüttner	-	-	-	-	1	-	-	-	-	-	-	-	-	-
Surirella minuta	Brébisson ex Kützing	-	-	-	-	1	-	-	-	-	-	-	-	-	-
Tabellaria flocculosa	(Roth) Kützing	-	-	-	-	-	-	-	-	-	1	-	-	-	1
Tabularia fasciculata	(C.Agardh) D.M.Williams & Round	-	1	-	-	-	-	-	-	-	-	-	-	-	-
Ulnaria ulna	(Nitzsch) Compère	-	-	-	-	1	1	-	-	-	-	1	1	-	-

Supplementary Figure 1. Cluster analysis of spring samples based on Jaccard similarity, illustrating patterns of resemblance in diatom species composition among the investigated springs.