Col. Vol. Sci. Pap. ICCAT, 83(1), SCRS/2026/026: 1-15 (2026)

GBYP BIOBANK: DEVELOPMENT OF BIOTUNA VISUALIZATION TOOL

I. Artetxe-Arrate[1], I. Zudaire[1], P. Lastra-Luque[1], N. Diaz-Arce[1], N. Rodriguez-Ezpeleta[1], F. Alemany[2] and I. Fraile[1]

SUMMARY

A central component of the ICCAT-GBYP programme is the systematic collection, preservation, and management of high-quality biological samples, stored in a long-term tissue bank. This repository currently contains thousands of otoliths, spines, gonads, and muscle or fin samples, along with associated analytical results such as stock-of-origin assignments and age estimates. To manage this expanding dataset, the BioTuna information system was developed as an interactive Shiny application that integrates metadata, sample availability, and analytical outputs. The platform provides spatial aggregation, biometric relationships, population-mixing estimates, and distributional summaries driven by user-defined filters. While visualizations are publicly accessible, metadata is released only through a regulated request process aligned with ICCAT data-protection rules. BioTuna enhances data interoperability, transparency, and scientific utility, maximizing the research value of the GBYP tissue bank under FAIR principles.

RÉSUMÉ

L'un des éléments centraux du programme ICCAT-GBYP est la collecte, la conservation et la gestion systématiques d'échantillons biologiques de haute qualité, conservés dans une banque de tissus à long terme. Cette base de données contient actuellement des milliers d'échantillons d'otolithes, d'épines, de gonades et de muscles ou de nageoires, ainsi que les résultats analytiques correspondants, tels que l'attribution des stocks d'origine et les estimations d'âge. Pour gérer cet ensemble de données en constante expansion, le système d'information BioTuna a été développé sous la forme d'une application Shiny interactive qui intègre les métadonnées, la disponibilité des échantillons et les résultats analytiques. La plateforme propose des concentrations spatiales, des relations biométriques, des estimations des mélanges entre populations et des résumés de la distribution, le tout sur la base de filtres définis par l'utilisateur. Si les visualisations sont accessibles au public, les métadonnées ne sont quant à elles communiquées que dans le cadre d'une procédure de demande réglementée, conforme aux règles de l'ICCAT en matière de protection des données. BioTuna améliore l'interopérabilité, la transparence et l'utilité scientifique des données, optimisant ainsi la valeur de la banque de tissus GBYP pour la recherche, conformément aux principes FAIR.

RESUMEN

Un componente fundamental del programa ICCAT-GBYP es la recogida, conservación y gestión sistemáticas de muestras biológicas de alta calidad, almacenadas en un banco de tejidos a largo plazo. Este repositorio contiene actualmente miles de muestras de otolitos, espinas, gónadas, músculo o aletas, junto con los resultados analíticos correspondientes, como la asignación a stocks de origen y las estimaciones de edad. Para gestionar este conjunto de datos en constante crecimiento, se desarrolló el sistema de información BioTuna, una aplicación interactiva de Shiny que integra metadatos, la disponibilidad de muestras y los resultados analíticos. La plataforma ofrece agregación espacial, relaciones biométricas, estimaciones de la mezcla de sstocks y resúmenes de distribución basados en filtros definidos por el usuario. Si bien las visualizaciones son de acceso público, los metadatos solo se facilitan mediante un proceso de solicitud regulado que se ajusta a las normas de protección de datos de ICCAT. BioTuna mejora la interoperabilidad, la transparencia y la utilidad científica de los datos, maximizando el valor de investigación del banco de tejidos del GBYP de acuerdo con los principios FAIR.

[1] AZTI, Marine Research, AZTI Basque Research and Alliance, Txatxarramendi ugartea z/g, 48395, Sukarrieta, Bizkaia, Spain
[2] ICCAT, Corazón de María 8, 28002, Madrid, Spain

KEYWORDS

GBYP, Biobank, Database, ABFT biology, Shiny, Biological sample repository

1. Introduction

The ICCAT-GBYP programme started in 2010, with the aim of improving understanding of Atlantic bluefin tuna (ABFT) key biological and ecological processes (https://www.iccat.int/gbyp/en/biostu.asp). Since then, one of the key activities of the programme has been to improve basic data collection, and the maintenance and management of a tissue bank for Atlantic bluefin tuna. This includes the preservation of high-quality biological samples that can be used for a wide range of bluefin tuna biology related scientific studies under the GBYP research plan. To date, the ICCAT-GBYP tissue bank has served as a repository of otolith, spines, gonads, muscle and/or fin tissues under rigorously controlled conditions to grant their long-term viability and scientific utility.

Presently, the tissue bank provides storage of 17,697 otoliths, 8,172 spines, 2,071 gonad samples and 24,266 muscle or fin tissue samples in its original or transformed form. The tissue bank also provides origin results of 795 fish analysed with the SNP array, origin results of 3,530 fish analysed with the 96 SNP panel, origin results of 2,679 fish analysed with otolith microchemistry, otolith derived age of 4,683 fish and spine derived age of 1,187 fish. The catalogue containing this information has been made available after each GBYP Biological studies phase, which was a global information system in a format of large Excel-based database, with 5 big modules of information: (1) Fish identification data, (2) Types of samples collected, (3) Sampling information, (4) Biological data and (5) Main results. However, the vast volume of information acquired over so many years highlighted that this system had important limitations: it did not link sample information with the analytical outputs and their results, and its structure made integration, exploration and interoperability difficult. These constraints underscored the need for a more robust and comprehensive solution and the creation of an advanced information system to streamline the search for data of interest and accessibility.

Since Phase 13 an online application has been developed for the visualization and management of biological sample availability and associated metadata. By consolidating diverse datasets into a unified, interactive platform, users can efficiently explore the composition of samples and visualize the outcomes of complex analyses. The tool has been designed to serve two key user groups: (1) Scientist from various institutions, who require a centralized catalogue of available samples and associated metadata to support diverse research activities, and (2) ICCAT members and fisheries managers, who benefit from a real-time and user-friendly interface to visualize and filter biological parameters essential for stock assessment and management advice.

2. Data and methods

Biotuna is an information system hosted in a public repository and delivered as an interactive Shiny application (http://aztidata.es/BioTuna). It provides a robust and reproducible framework for visualizing and managing biological sample availability and associated metadata within the GBYP tissue bank, among others. The interface lets users dynamically choose parameters such as geographic areas, time windows, sample types, biological characteristics etc., and explore them through faceted filters and responsive maps, tables and plots.

Currently the interface contains a search engine and four tabs; Map, Data distribution, Results and Data.

2.1 Search engine

The search engine counts with 12 filters that can be used in combination in order to project the desired search in the display tabs (**Table 1**). Filters are selected by default, and users can refine their search by applying specific criteria. After selecting filters, clicking the "*Apply*" button updates the interactive elements, while the "*Clean*" button resets the search to start over again.

Table 1. List of filters and their descriptors available for users to perform targeted searches.

FILTER NAME	*DESCRIPTION*
Project	Indicates the research project or funding initiative under which the samples were originally collected and catalogued. For ICCAT-GBYP related samples please click on "Deselect All" and then select only "GBYP".
Species	Multiple choice selector for species by scientific name. For Atlantic bluefin tuna samples please click on "Deselect All" and then select only "Thunnus thynnus"
FAO-Area	Multiple choice selector for areas as defined by FAO major fishing areas for statistical purposes.
Sub-Area	Multiple choice selector for smaller areas defined inside the FAO areas and defined as sampling strata used for GBYP Biological Studies
Catch year	Multiple choice selector for the year when the fish was caught from the wild.
Birth year	Multiple choice selector specific cohorts (estimated from age-length relationship).
Size class	Multiple choice selector for the class assigned according to fish weight.
Straight fork Length	The bar allows selecting the desired size range.
Total Weight	The bar allows selecting the desired weight range.
Length based age	The bar allows selecting the desired age range, derived from length-based age.
Sex	Multiple choice selector for sex.
Sample type	Multiple choice selector for each sample type (otolith, muscle, fin, spine, gonad and stomach). There is an additional selector named "AND/OR (cheked: AND), when this selector is not checked, tabs will show results for fish containing any of the samples types selected. When this selector is activated, tabs will show only those fish containing all the selected sample types.

2.2 Map tab: Visualizing geographic sample distribution

The "Map" tab of the Shiny application provides a global overview of sample distribution based on the filters selected by the user (see 2.1 Search engine). In order to ensure catch position privacy while maintaining the utility of spatial data visualization, the "Map" tab aggregates sample availability by FAO major fishing areas and sub-areas, using a blue-scale color-coded system to represent the relative abundance of samples within each sub-area (**Figure 1**). This approach preserves the anonymity of exact sampling locations while providing meaningful information on geographical coverage of ABFT samples and individuals.

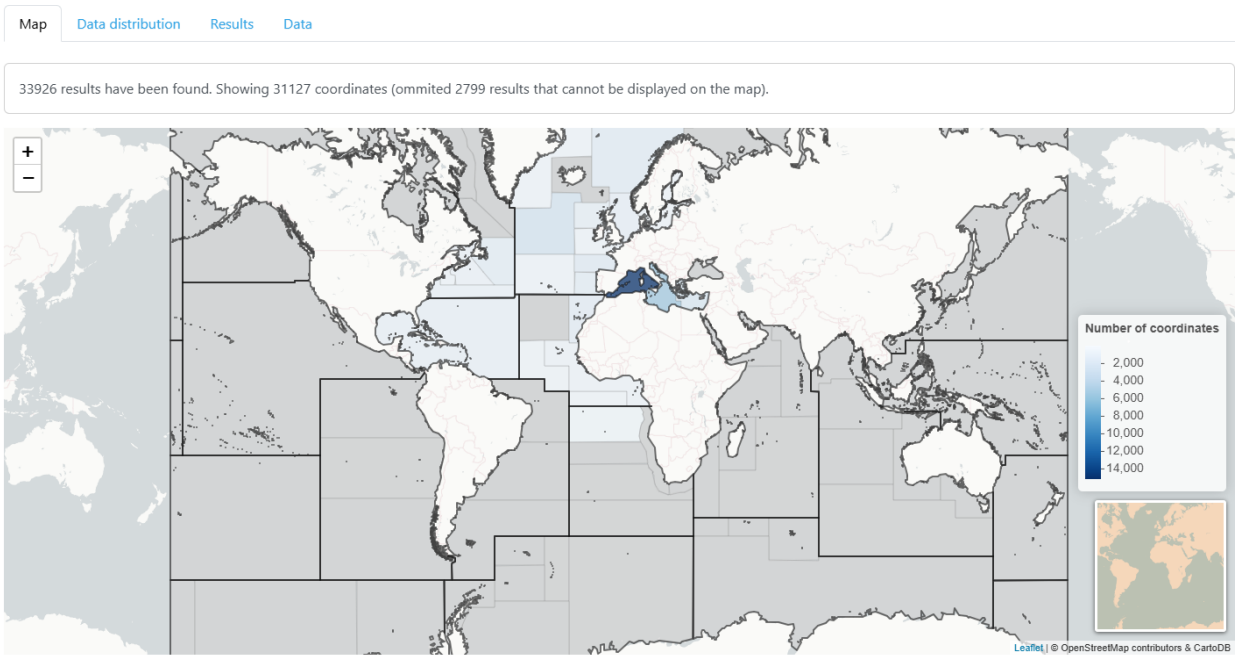


Figure 1. Map tab of the Biotuna shiny application, showing geographic distribution of the ABFT data collected under GBYP Biological sampling programme. Blue-scale colour intensity represents relative sample abundance within each sub-area, ensuring confidentiality while providing an informative spatial overview. Number of coordinates in the legend refers to the number of fish with geospatial data that meet the criteria selected of the "Filters" section.

In addition to the general overview, the map also supports zooming into predefined sub-areas, sampling strata pre-defined within FAO regions, offering a more detailed view of sample distribution without compromising confidentiality (**Figure 2**). The map is fully interactive and responds dynamically to the filters applied on the left-hand panel (see 2.1 Search engine), which includes, catch year, size class, sample type, and other biological or geographic criteria.

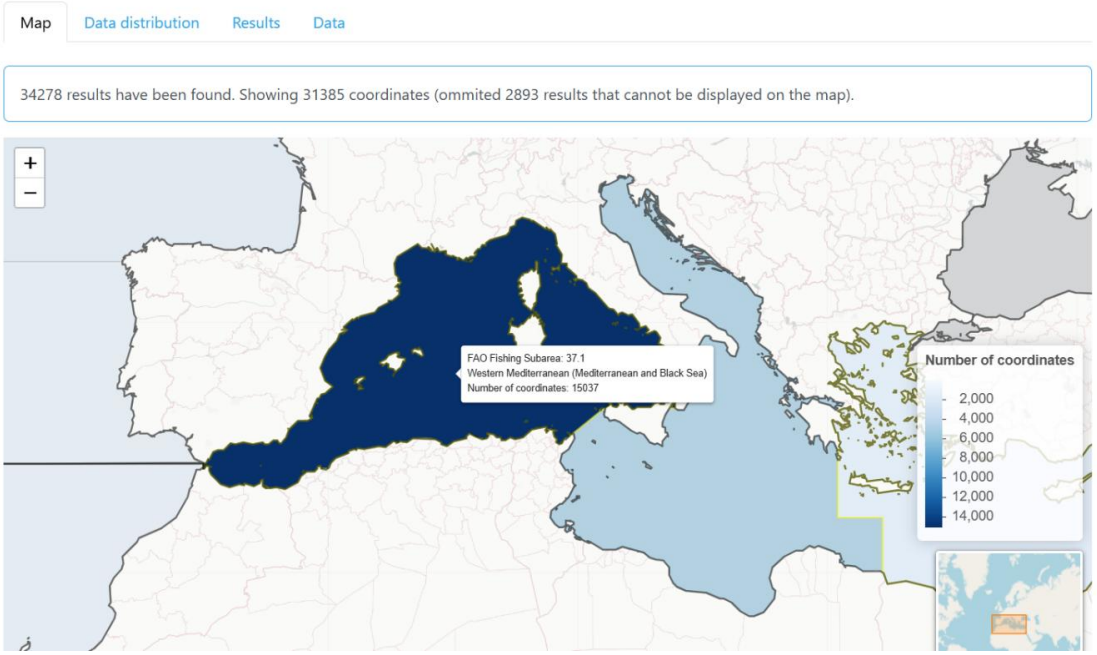


Figure 2. Zoomed view of the Map, illustrating sample availability aggregated by FAO fishing sub-areas. In this example, the Western Mediterranean Sea (FAO Subarea 37.1) is highlighted, showing the number of coordinates (number of fish sampled) that meet the selected filter criteria.

2.3 Data distribution tab

The "Data distribution" tab a set of interactive visualizations intended to facilitate empirical examination of the dataset composition and facilitate data-driven decision-making. This functionality enables users to explore the structure and characteristics of the available fish through five distinct distribution bar charts:

Length Distribution: Displays the frequency of samples across different straight fork length ranges, offering insights into size composition (**Figure 3**).

Weight Distribution: Illustrates the distribution of total weight, enabling users to identify patterns in biomass representation (**Figure 4**).

Age Distribution: Shows the estimated age of individuals based on length-age relationships, which is critical for life history and growth studies (**Figure 5**).

Distribution by Year of Capture: Highlights temporal trends in sampling effort and availability, useful for understanding interannual variability (**Figure 6**).

Distribution by Year of Birth: Provides cohort-based visualization, supporting analyses of recruitment and population dynamics (**Figure 7**).

Each chart is fully interactive and responds to the filters applied. Users can adjust bin sizes to refine the granularity of the visualization, making it easier to detect patterns or anomalies (e.g. **Figure 4**). The data can be viewed in the bars globally, if the 'All' selector is clicked (e.g. **Figure 3** and **4**), by Sex if the 'Sex' filter is clicked (e.g. **Figures 5** and **6**), or by FAO area if 'FAO Area' is clicked (**Figure 7**). Additionally, if specific levels wanted to be viewed within sex and FAO area, they can be directly selected or deselected on the image itself (e.g. **Figure 6**).

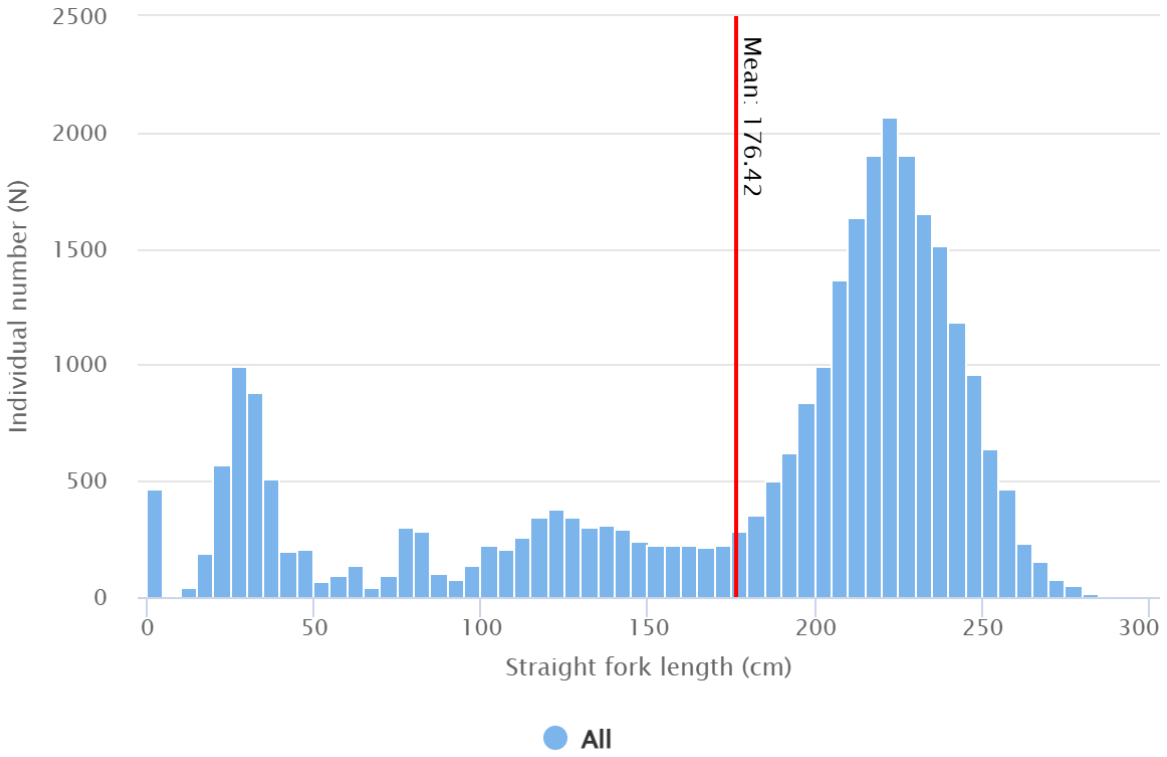


Figure 3. Straight fork length (cm) data distribution showed by 5 cm bin size. Red vertical line shows mean distribution.

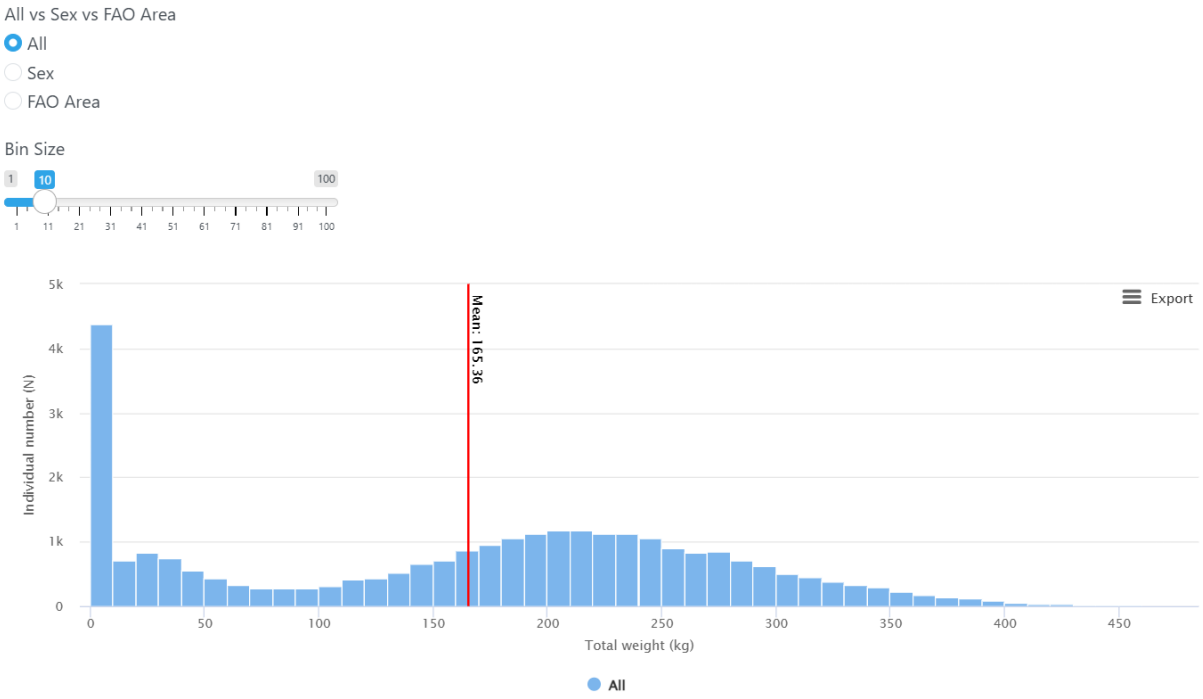


Figure 4. Total weight (kg) data distribution by 10 kg bin size. Red vertical line shows mean distribution.

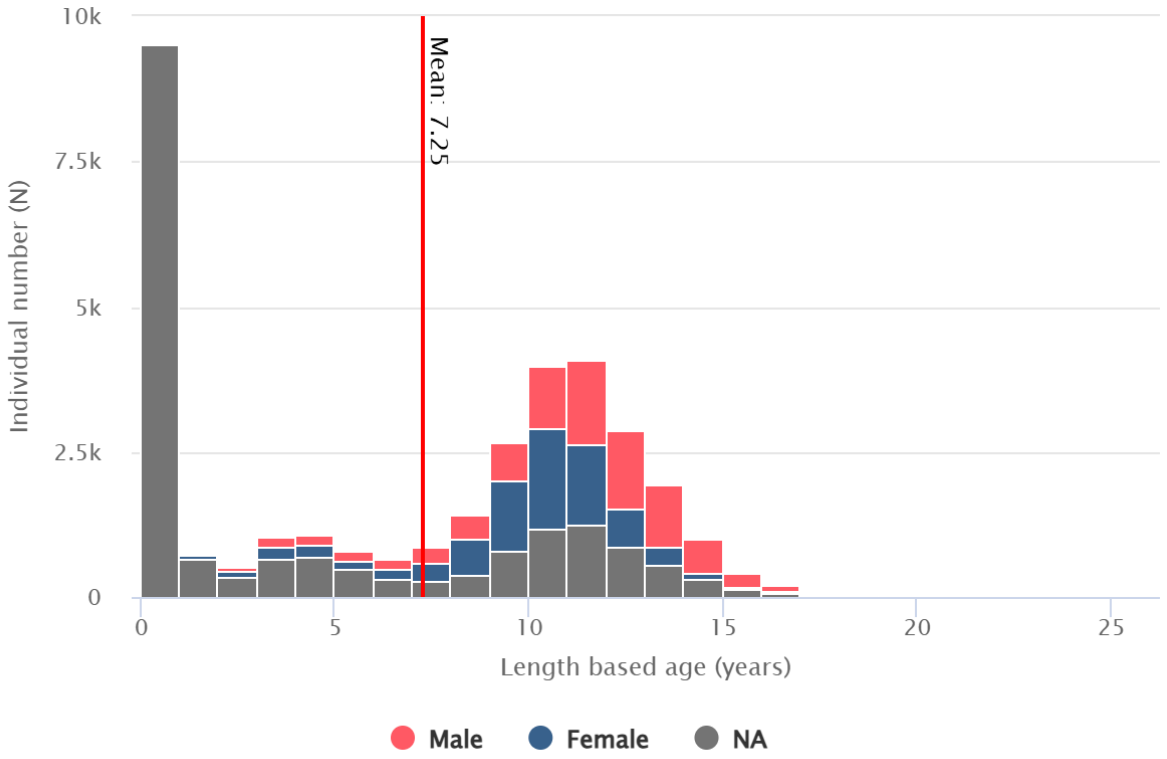


Figure 5. Length based age (years) data distribution, where proportions are showed by sex (male, female or NA). Red vertical line shows mean distribution.

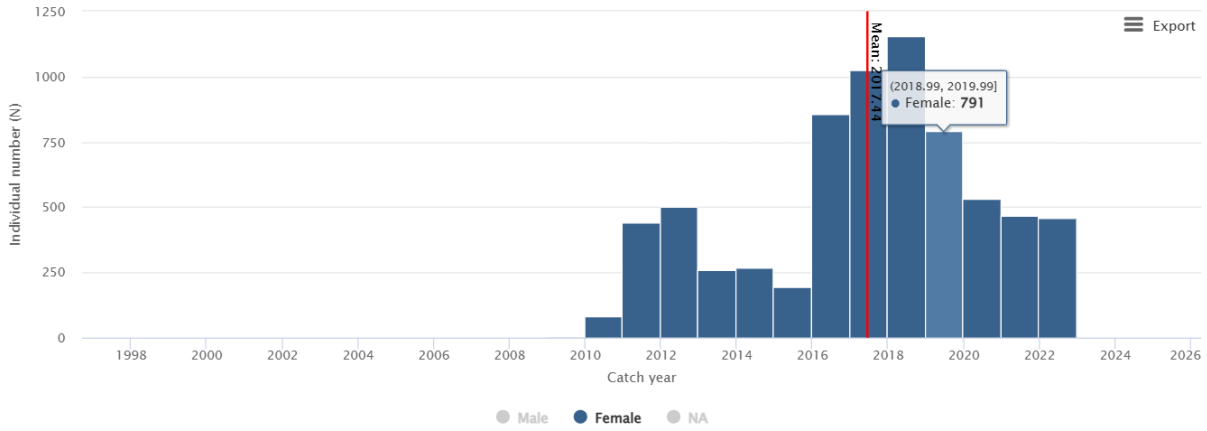


Figure 6. Catch year data distribution, where data is only showed for females. Red vertical line shows mean distribution. When clicking on an specific bar, number of females is shown.

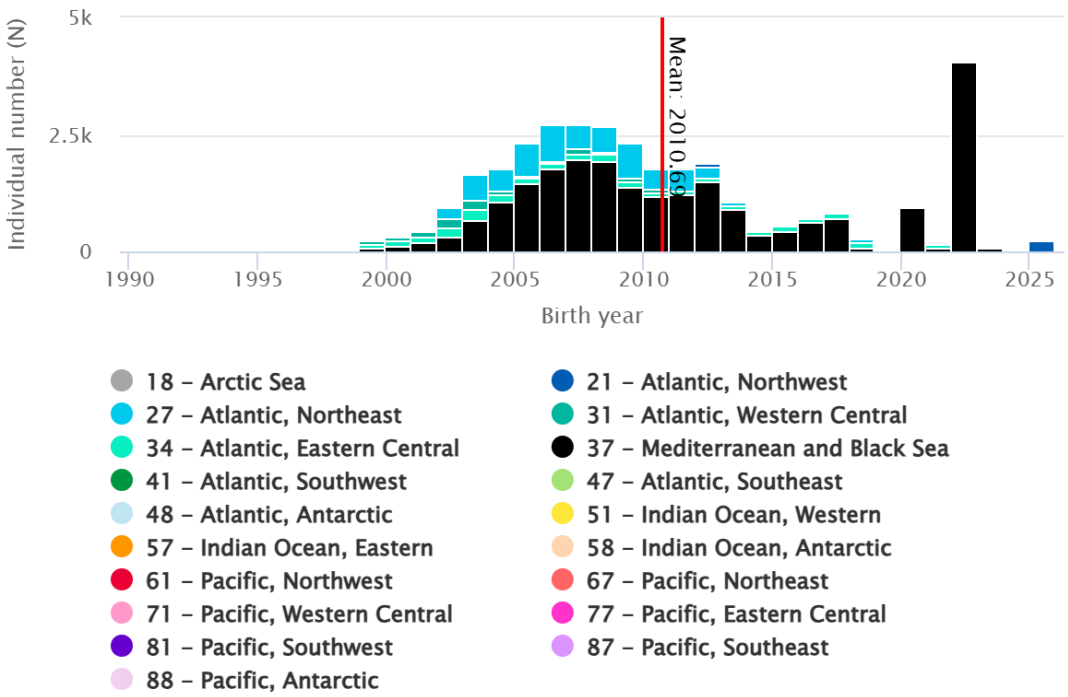


Figure 7. Birth year (estimated from age-lengh keys) data distribution by FAO Area. Red vertical line shows mean distribution.

2.4 Results tab

The "Results" tab offers real-time visualization of biometric relationships, including: (1) fish length–weight, (2) fish length–otolith weight, and (3) fish weight–otolith weight. These relationships and their equations are dynamically calculated according to the selected filters, ensuring that the outputs reflect the subset of data relevant to the user's query. Each plot is accompanied by the corresponding regression equation and the coefficient of determination (R^2), providing an indication of the model fit (**Figure 8** and **9**). These relationships are essential for growth studies, age estimation, and stock assessment modelling, as they allow researchers and managers to derive accurate conversions between key biological parameters.

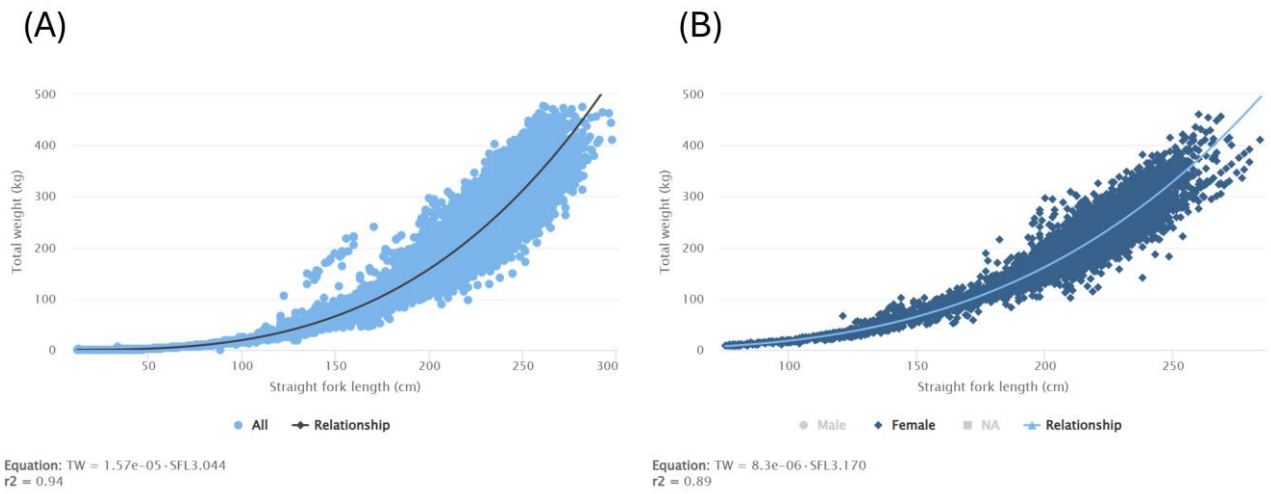
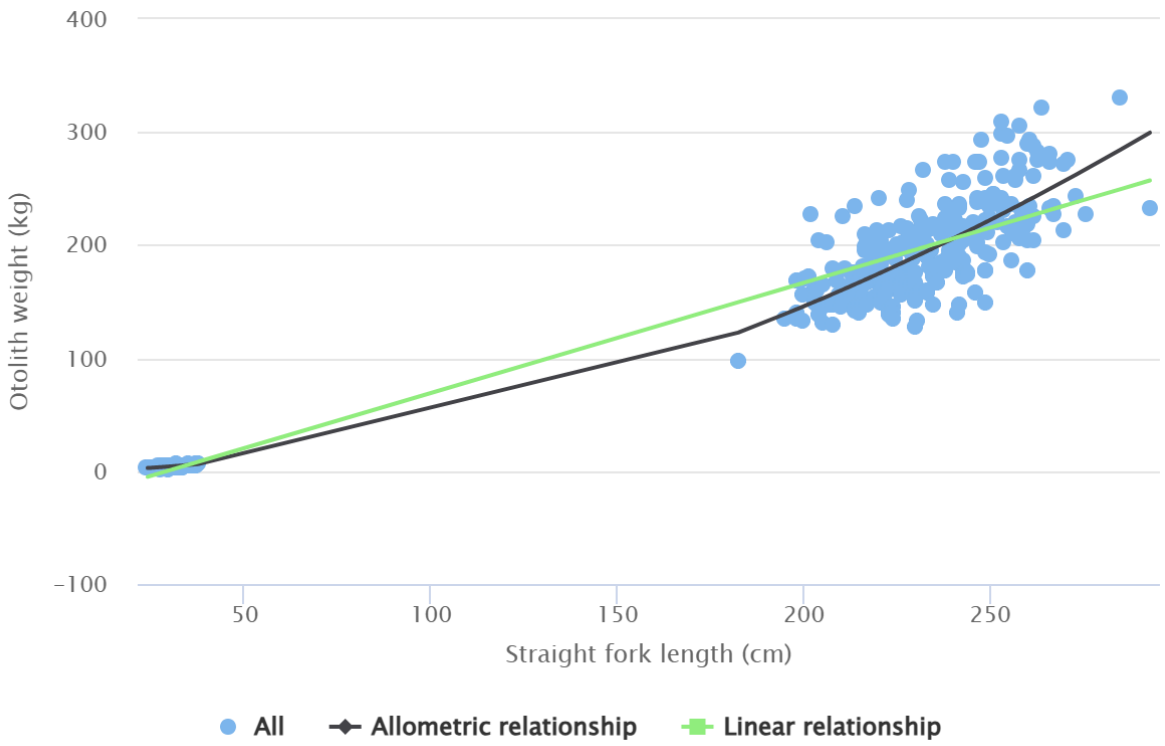


Figure 8. Example of length-weight biometric relationship graphs displayed in the "Results" tab; (A) All ABFT lengh-weigth relationship and (B) length-weight relationship filtered for female ABFT only. Relationship equation and R^2 are shown for each graph.



Allometric relationship:
Equation: OW = 6.6485e-03 · SFL1.887
r2 = 0.92
Linear relationship:
Equation: OW = -28.392 + 0.975 · SFL
r2 = 0.91

Figure 9. Example of fish length-otolith weight relationship graphs showhing two types of relatinships and equations; allometric relationship (in black) and linear relationships (in green).

Additionally, the 'Results' tab includes graphical representations of sex ratio (**Figure 10**) and maturity curves (**Figure 11**). As with other visualizations, these charts are automatically updated according to the chosen filters (e.g., species, FAO area, size class, catch year), ensuring relevance to the user's query.

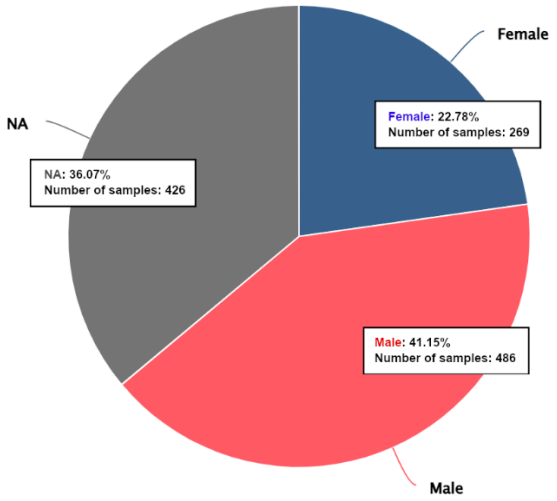


Figure 10. Sex ratio distribution within the filtered dataset, showing the proportion of males (41.15%), females (22.78%), and individuals with no available sex information (NA: 36.07%). The interactive pie chart allows users to hover over each segment to view exact percentages and sample counts.

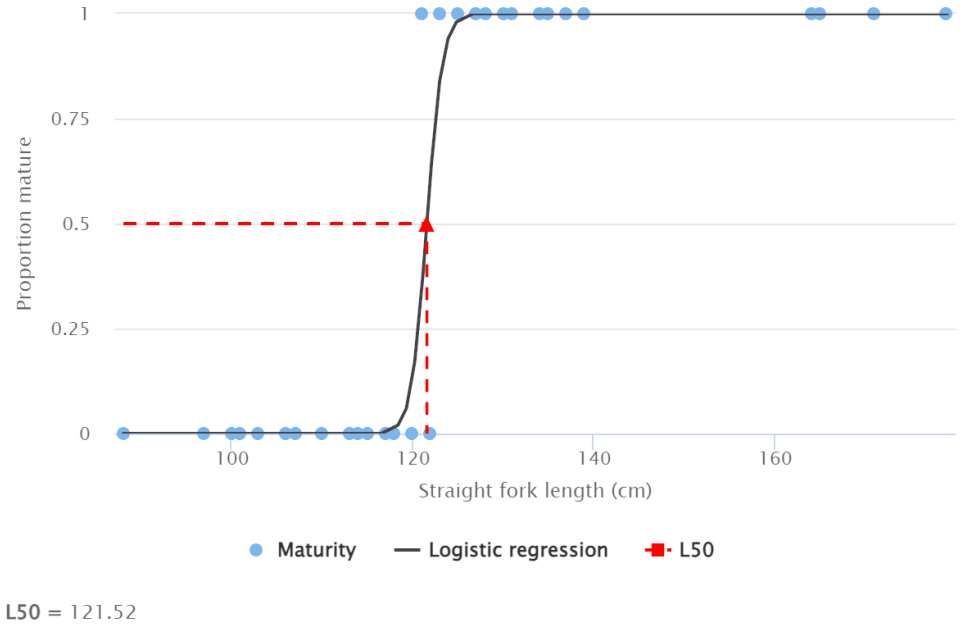


Figure 11. Maturity curve for bluefin tuna (*Thunnus thynnus*) based on logistic regression. The plot shows the proportion of mature individuals as a function of straight fork length (cm). Blue points represent observed maturity data, while the black line indicates the fitted logistic regression curve. The red dashed lines highlight the length at 50% maturity (L_{50} = 121.52 cm), which is a key parameter for reproductive biology and stock assessment.

Finally, this section provides an option to visualize population mixing proportions (Mediterranean vs. Gulf of Mexico) for Atlantic bluefin tuna, derived from otolith microchemistry analyses, the 96-SNP panel, and the SNP array (**Figure 12**). These estimates have been calculated following the latest criteria adopted within the GBYP Biological Studies, ensuring consistency with current best practices. For transparency and reproducibility, the details of the methodology used to generate these estimates are fully described in the latest GBYP report (Fraile et al. 2025).

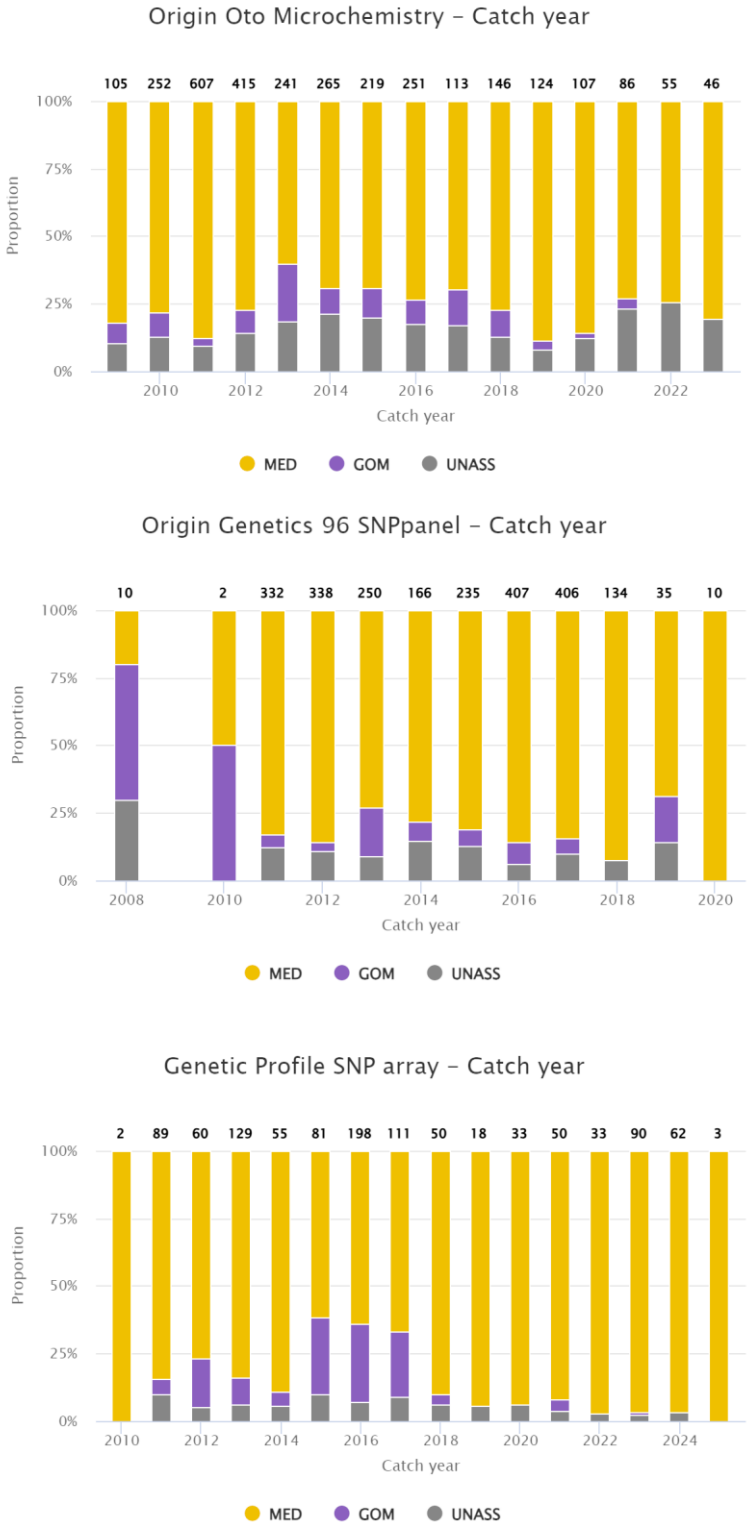


Figure 12. Example of stock of origin by catch year based on three analytical approaches. The figure compares annual proportions of Atlantic bluefin tuna assigned to three origin categories: Mediterranean (MED), Gulf of Mexico (GOM), and unassigned (UNASS), using otolith microchemistry (upper panel), a 96-SNP genetic panel (central panel), and a high-density SNP array (lower panel). Numbers above each bar indicate the sample size analyzed per year.

To facilitate reporting and sharing, the application also offers the option to export any chart in multiple formats, including PNG, JPG, PDF, and SVG. This feature allows users to easily incorporate visual outputs into presentations, reports, or scientific publications in real time without additional processing.

2.5 Data tab: metadata, data and/or samples requesting

All visual outputs generated by the application, such as maps, charts, and output graphics, are publicly accessible, and can be downloaded from the online application. However, access to the underlying metadata is restricted. Users wishing to obtain metadata must complete an online request form by selecting "Request access", which enables the GBYP tissue bank coordinators to evaluate the purpose of the request and grant time-limited access in accordance with ICCAT data-protection procedures (https://iccat.int/Data/Data_protection_ENG.pdf).

To access metadata, the online request form that users must complete **(Figure 13**). The form collects essential information needed to evaluate the metadata request, including:

Personal Information: Full name, email, country, telephone, and institution/organization.
Project Details: Project title, main objective, brief description (max. 200 characters), and project duration.
Requested Data: Sampling years, areas, species, and intended use (e.g., research, education, other).
Publication Intent: Indication of whether metadata will be used in publications.
Declaration: Confirmation that the data will only be used for the specified purpose.

Once submitted, the request is reviewed by the ICCAT-GBYP tissue bank coordinators who will contact the GBYP programme coordinator to share the request. Approval of the request follows the guidelines of the "*Rules and Procedures for the protection, access to, and dissemination of data compiled by ICCAT*" document, specifically rules described in paragraph 9. (bis), which classifies biological data as medium-risk, non-public data. Consequently, metadata provided will never include confidential information or exact fishing coordinates. If approved, the user receives temporary, password-protected access to the requested metadata. Access is limited to the purpose and duration specified in the request, ensuring responsible use and full traceability. Users are also informed that the data source must be acknowledged in any resulting outputs.

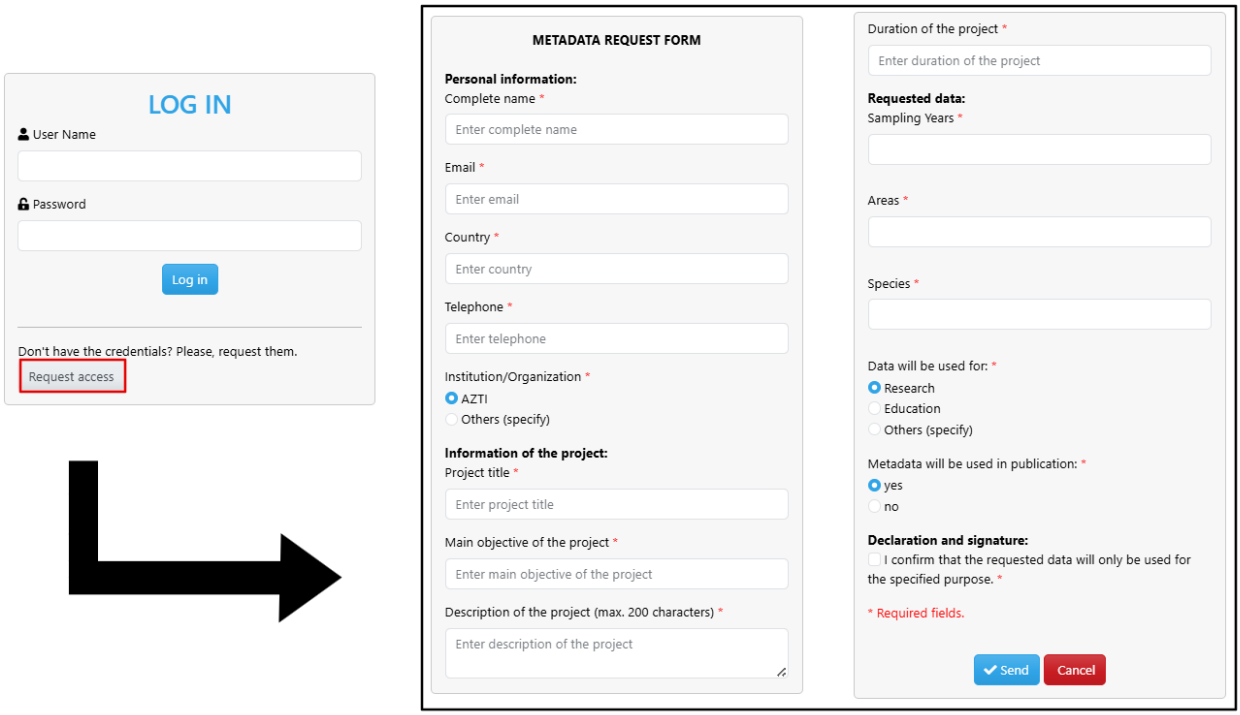


Figure 13. Access to metadata request in the BioTuna platform. The left panel shows the login interface for users with existing credentials. For new users, clicking "Request access" opens the metadata request form (right panel), where personal details, project information, requested data, and intended use must be specified. Upon approval by the platform coordinators, temporary credentials are provided for accessing the requested metadata.

However, while metadata can be accessed directly and, when appropriate, made openly available, their access should be managed by the BioBank coordinators, who will inform the ICCAT Secretariat/GBYP to ensure proper tracking and reporting to the SCRS and the Commission on the effective use of the biobank and its associated information system. In contrast, any request for raw data or physical samples—at least those belonging to GBYP—must be submitted directly to the ICCAT Secretariat through the official procedures and forms established in the "Rules and Procedures for the protection, access to, and dissemination of data compiled by ICCAT" after which the request will be evaluated by the committee defined in those same rules (https://iccat.int/Data/Data_protection_ENG.pdf).

3. Results and conclusions

By bridging the gap between raw data and applied research, the tool supports informed decision-making, promotes data sharing, and enhances the visibility and impact of the work carried out within the ICCAT-GBYP project. The information system provided via a shiny page interface enhances accessibility and transparency, enabling dynamic data visualization, improves research planning and informed decision-making, strengthens transparency, and promotes collaboration among institutions. This approach not only accelerates research progress but also ensures that the tissue bank serves as a robust resource for advancing knowledge in ABFT biology and dynamics. Moreover, the tool plays a critical role in maximizing the scientific value of the GBYP tissue-bank by making **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable (FAIR principles) (Wilkinson *et al.,* 2016). These principles provide a foundation for improving data stewardship and for enhancing the long-term scientific value of biological samples and associated analytical outputs. The continued development of the information system is necessary to incorporate new samples and results, ensure metadata consistency, and adapt the system to evolving analytical needs and data standards. This ensure that information system becomes a central resource for data sharing and knowledge dissemination within the ICCAT GBYP programme.

Acknowledgements

This work was carried out under the provision of the ICCAT Atlantic-Wide Research Programme for Bluefin Tuna. The contents of this document do not necessarily reflect the point of view of ICCAT, which has no responsibility over them, and in no ways anticipate the Commission's future policy in this area. The current version of the shiny app was co-funded by the European Union, partially funded through the EU Grant Agreement No.101169569-GBYP Phase 14 project and the European Climate, Infrastructure and Environment Executive Agency (CINEA) Gran Agreement No.101157296 -ITUNNES project.

References

Fraile, I., Artetxe-Arrate, I., Diaz-Arce, N., RodriguezEzpeleta, N., Lastra-Luque, L., Zudaire, I., Etxebarria, S., Gutierrez, N., Mendibil, I., Orbe, A., Garcia, G., and Serrano, N. (2025). Draft final report on short term contract for biological studies - ICCAT Atlantic-Wide Research Programme for bluefin tuna (ICCAT-GBYP Phase 14-2025).

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, *3*(1), 1-9.