

**Specimen and Material
Citation Matching
Services**

What is the eBioDiv Matching Service?

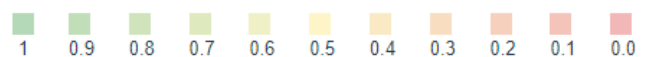
The Earth’s scholarly knowledge about biodiversity is included in a corpus of several hundred million pages of academic publications spanning over 250 years. These publications often contain references to biological specimens in natural history collections. In a digital world, access to this trove of biodiversity knowledge would be greatly enhanced if the references between the scholarly articles and the cited specimens were bi-directional and machine readable. The eBioDiv Matching Service helps bridge this gap. The crowdsourcing tool allows users to match material citations contained

in the academic literature to the respective specimens in natural history collections.

How does it work?

To facilitate the task, a semi-automatic approach is used, where users are presented with lists of possible matches, along with matching scores indicating the probability of a match calculated by an algorithm developed for this purpose. Over time, the matching decisions taken by the users will be used as input to further refine the algorithm, thus increasing the efficiency of the tool.

Color legend for the matching score



Specimen 2513789592 from Naturalis Biodiversity Center

Locality	Country	Date	Institution code	Collection code	Catalog nb	Individual nb	Collector (recorded by)	Type	Record
Vogelkop Peninsula, Ije river valley, Central Tamrau Ra., S. slope, Mt. Kusemun, ascent from Sudjak village.	Indonesia	8/11/1961		Botany	L.1493016		Royen P van; Sleumer HO; Schram FAW	Holotype	Specimen

Material citations associated with the specimen 2513789592

1 suggested material citation to curate

Curate for each suggested material citation whether it matches the specimen or not (Yes / No).

Locality	Country	Date	Institution code	Collection code	Catalog nb	Individual nb	Collector (recorded by)	Type	Record	Decision
Sudjak village	Indonesia	8/11/1961		S		1	Van Royen & Sleumer & Schram	Holotype	Material citation	<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Undecidable



Where does the data come from?

The eBioDiv Matching Service processes data that has been published through the Global Biodiversity Information Facility (GBIF). As the service is used to match material citations from the academic literature with the related specimens in natural history collections, it relies on intermediary data sources of two types: 1) Material citations databases; 2) Collection databases (from institutions).

Material citations databases

They are used to store the information from taxonomic treatments found in the academic literature in the form of structured data. As regards material citations data, the eBioDiv Matching Service mainly relies on the data provided through the Plazi TreatmentBank, which is currently the world's largest material citations database. Since its origins, the data provided through TreatmentBank has been extracted from legacy sources (e.g. scans of academic journals published in the past) or born-digital PDFs with the help of computer algorithms. The Plazi TreatmentBank is regularly harvested by GBIF. Thus, all the material citations contained in TreatmentBank are also available on GBIF and serve as inputs to the eBioDiv Matching Service.

Species: *Bulbophyllum cerastes* J.J.Verm., Schuit. & de Vogel
Location: Indonesia
Elevation: 1100m
Basis of record: Material citation ←
Specimen type: Holotype

Collection databases

They are used to store the metadata about the individual items of a collection. When it comes to “digitising” a natural history collection, one of the first steps consists in creating a digital catalogue, or digital inventory, of all the items in the collection. When it comes to sharing collection metadata beyond institutional boundaries, the “digitization” process typically also involves adhering to a system of globally unique identifiers as well as agreeing upon and adhering to a harmonised data model, and ingesting it into a common database that serves as a data aggregator. One such data aggregator at the international level is GBIF.

Species: *Bulbophyllum cerastes* J.J.Verm., Schuit. & de Vogel
Location: Asia > Indonesia
Basis of record: Preserved specimen ←
Specimen type: Holotype

