

The challenges of integrating biodiversity datasets: Can taxonomic databases help?

SMARTSkills workshop

Galway, October 24-25 2013

What is biodiversity?

Biodiversity= number of species + their abundance



The diagram features a light blue rounded rectangular box containing the text 'Biodiversity= number of species + their abundance'. A red arrow points from the text 'Taxonomic data' below to the words 'number of species' inside the box. Another red arrow points from the text 'Numerical data' above to the words 'their abundance' inside the box.

Taxonomic data

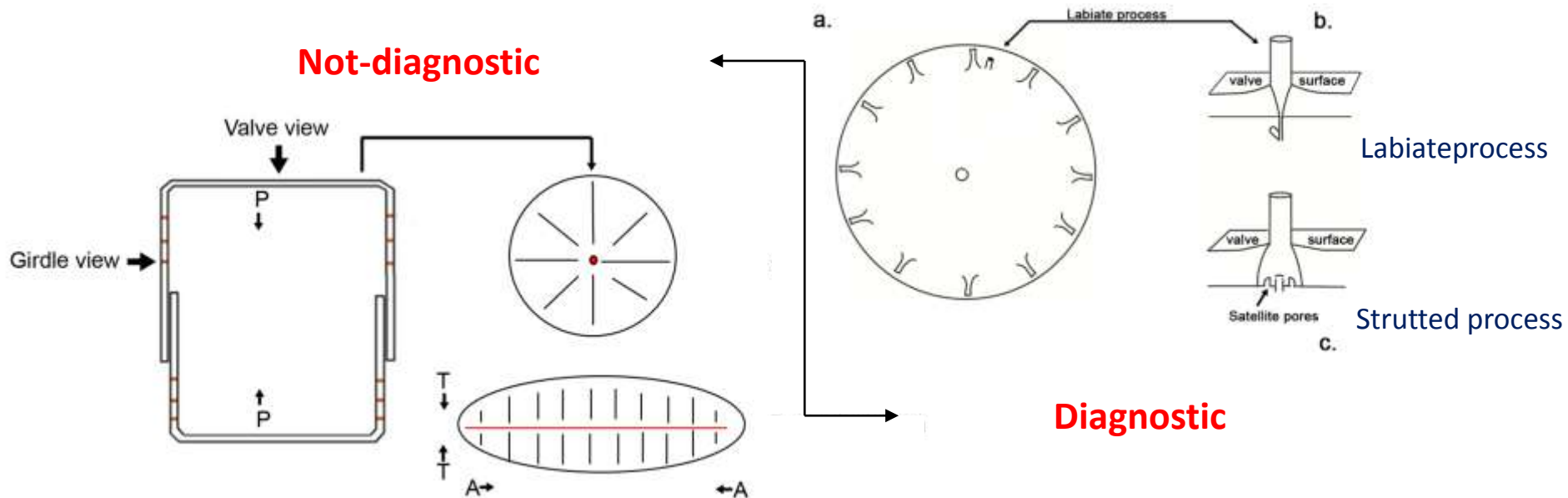
Numerical data

- Complex metadata
- Need to link different types of data
- Difficult to compare between datasets (taxonomic/identification problems)

Taxonomic data

- Taxa are identified based on morphological and increasingly molecular data
 - Diagnostic value of a character has to be established.
 - Species are classified according to the similarity (relatedness) between morphological characters

Example: Diatom morphology



With strutted process=Thalassiosiraceae, without it= Coscinodisceae and others

Structure of taxonomic data

Taxonomic data have a hierarchical structure (metadata)

Example: *Thalassiosira*

Kingdom: Chromista

Phylum: Heterokontophyta/Ochrophyta

Class: Coscinodiscophyceae

Order: Thalassiosirales

Family: Thalassiosiraceae

Genus species: *Thalassiosira*



Live cells of *Thalassiosira punctigera*

Problem: The interpretation of characteristics can change

Synonyms:

Ethmodiscus punctiger Castracane 1886, *Coscinodiscus punctiger* (Castracane) H.Peragallo 1889
Ethmodiscus japonicus Castracane, *Coscinodiscus angstii* Gran 1931, *Thalassiosira japonica*
Kisselev 1935, *Thalassiosira angstii* (Gran) Makarova 1971

Integrating biodiversity datasets:

- Need to take into account taxonomic metadata, naming conventions etc.

→ Taxonomists/analysts rarely agree

Dataset 1

Kingdom: Chromista

Phylum: ~~Heterokontophyta~~/Ochrophyta

Class: Coscinodiscophyceae

Order: Thalassiosirales

Family: Thalassiosiraceae

Genus species: *Thalassiosira punctigera*

Dataset 2

Kingdom: Chromista

Phylum: Heterokontophyta/~~Ochrophyta~~

Class: Coscinodiscophyceae

Order: Thalassiosirales

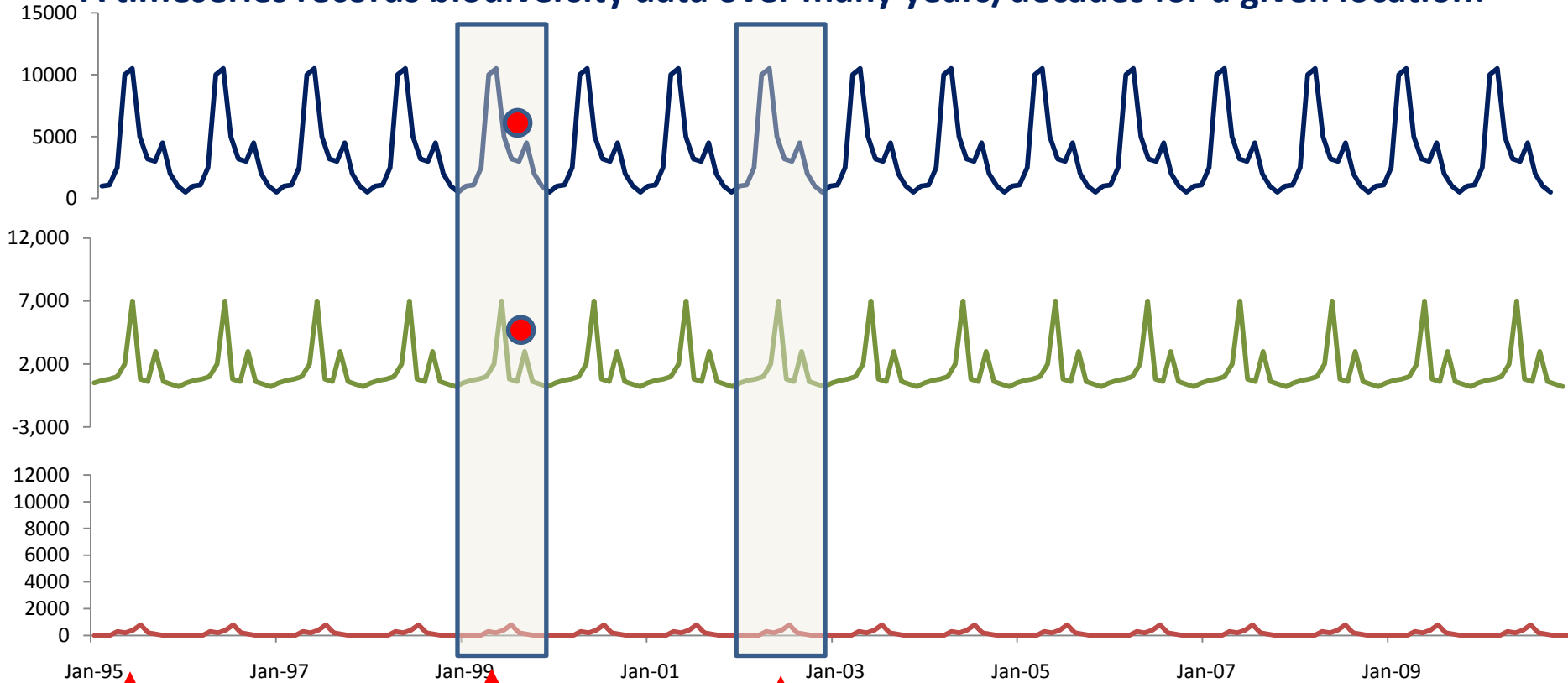
Family: Thalassiosiraceae

Genus species: *Thalassiosira punctigera*

Additionally: a lot of unidentified species, recorded as size classes

Implications for timeseries data

A timeseries records biodiversity data over many years/decades for a given location:



● =Taxon re-named

Total number of species recorded per year

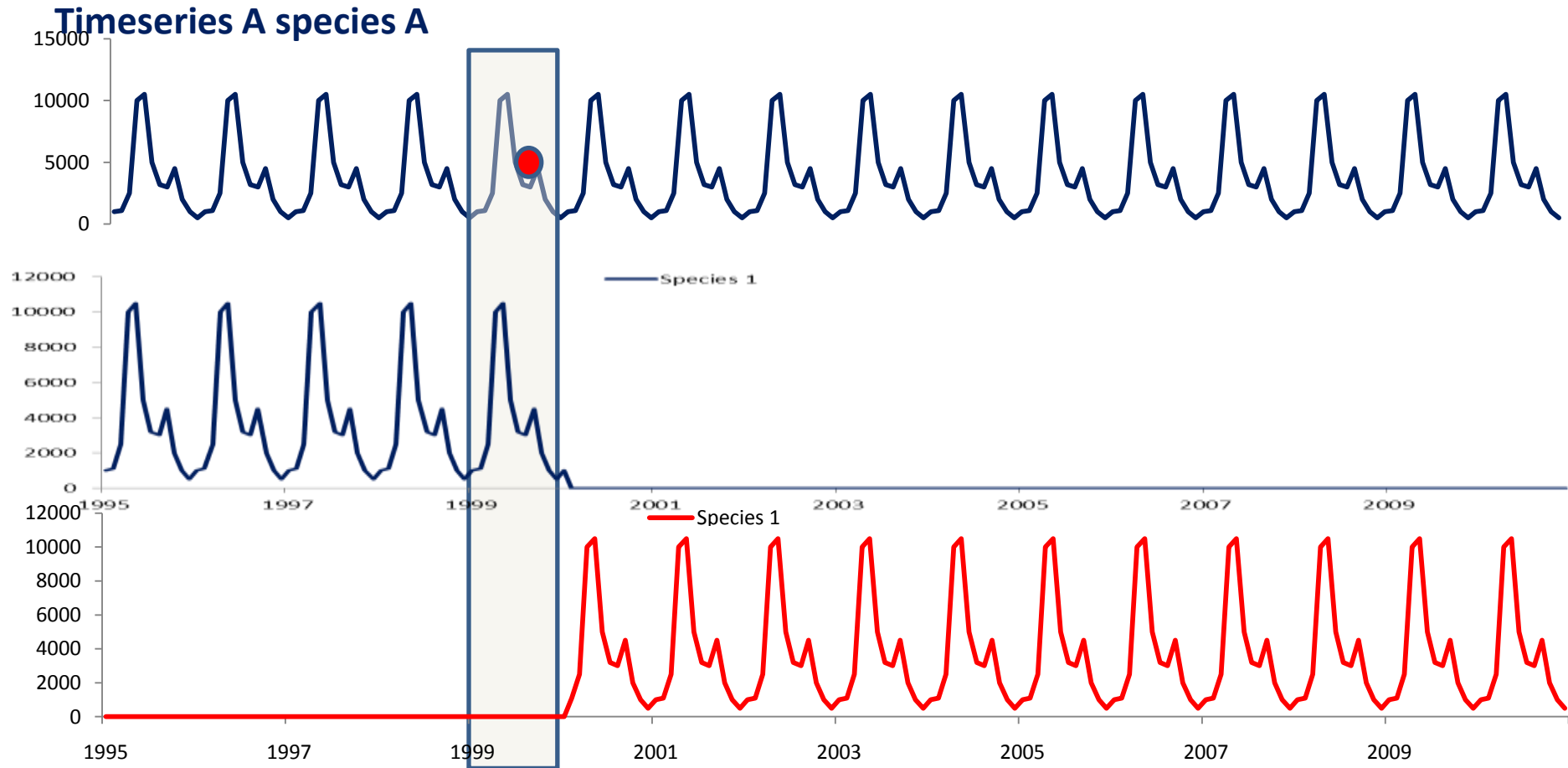
3

5

3

Different names

Artifacts within timeseries

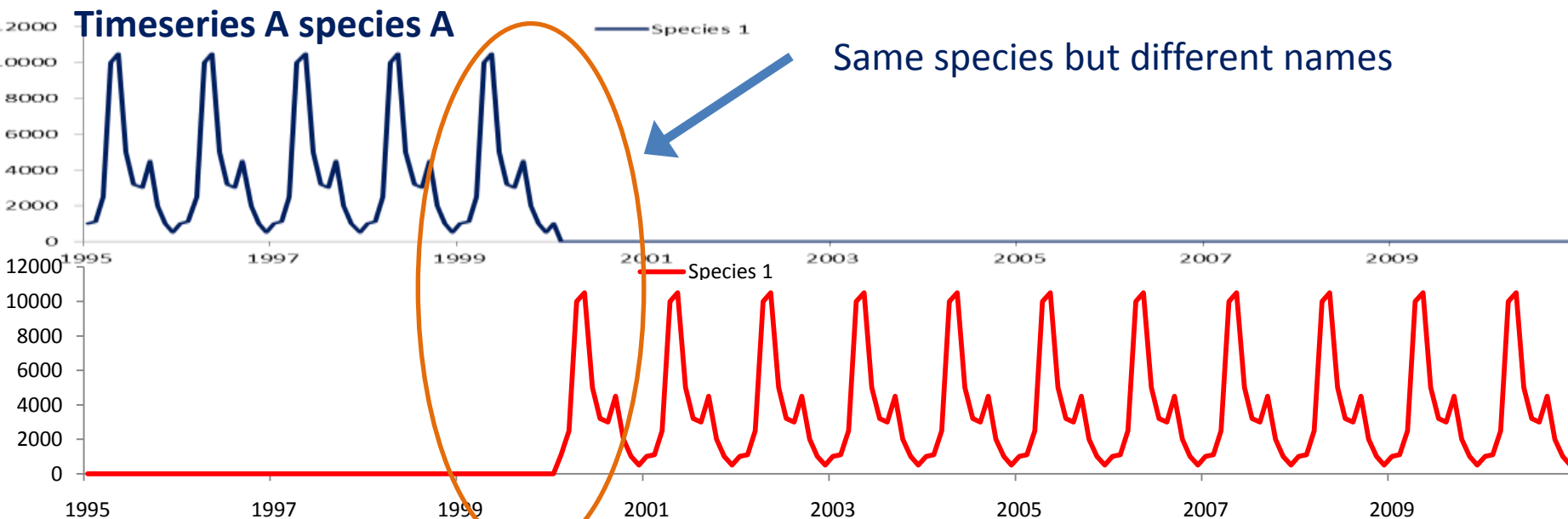


Name change documented in one data set but not in the other:

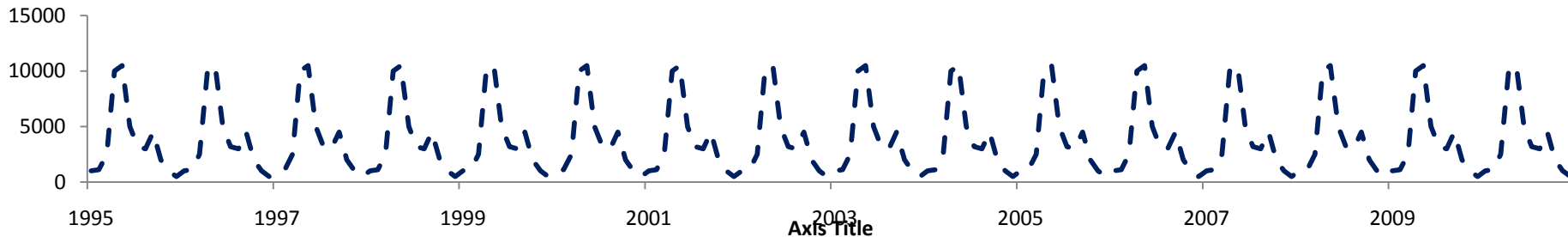
- Datasets not comparable
- Need intercalibration exercise

 = Taxon re-named

Comparison between timeseries



Timeseries B species A



Structure of taxonomic data

Taxonomic data have a hierarchical structure (metadata)

Example: *Thalassiosira*

Kingdom: Chromista

Phylum: Heterokontophyta/Ochrophyta

Class: Coscinodiscophyceae

Order: Thalassiosirales

Family: Thalassiosiraceae

~~Genus species: *Thalassiosira punctigera*~~



Live cells of *Thalassiosira punctigera*

Problem: The interpretation of characteristics can change

Synonyms:

Ethmodiscus punctiger Castracane 1886, *Coscinodiscus punctiger* (Castracane) H.Peragallo 1889
Ethmodiscus japonicus Castracane, *Coscinodiscus angstii* Gran 1931, *Thalassiosira japonica*
Kisselev 1935, *Thalassiosira angstii* (Gran) Makarova 1971

Difficult data discovery

Data are not static: Many synonyms per species

Example: *Neocalyptrella robusta* (1) (*Calyptrella robusta* (6), *Rhizosolenia robusta* (1342))



Standort - Google Analy... Umfrage: Deutsche Haber... biodiversity data provider... Algalbase: Listing the W... Data Publisher for Earth &...

pangaea.de/search?count=10&minlat=&minlon=&maxlat=&maxlon=&mindate=&maxdate=&env=All&q=thalassionemataceae

Google-Ergebnis für... Vorgeschlagene Sites Web Slice-Katalog physiology_stuff Marine Microbial Eu... Pilates: Methode, Ü... Übung: Phönix - Pila... Weitere Lesezeichen

All Water Sediment Ice Atmosphere

thalassionemataceae Search

Help Advanced Search Preferences more...

Not logged in (log in or sign up)

Always quote citation when using data!

0 datasets found on search for thalassionematac... Show Map Google Earth Data Warehouse

<< PREV | NEXT >>

<< PREV | NEXT >>

Contact

- Difficult to find data for a taxon under all names
- Difficult to aggregate data for higher taxonomic levels
- Ideally timeseries should be archived together with images (as metadata)

→ Need databases with a taxonomic backbone and synonym management


Plankton data

- Added problem: specimen are not visible with the naked eye
- Not easy to preserve samples
 - need to have photographic documentation
 - Needs to be documented with ist own metadata


PLANKTON*NET

(<http://planktonnet.awi.de>)

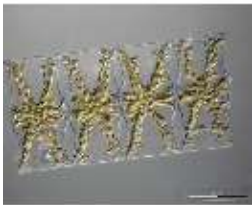
- Presence/absence datasets
- metadata for numerical data
- data are georeferenced
- 10000 images archived
- classification, references



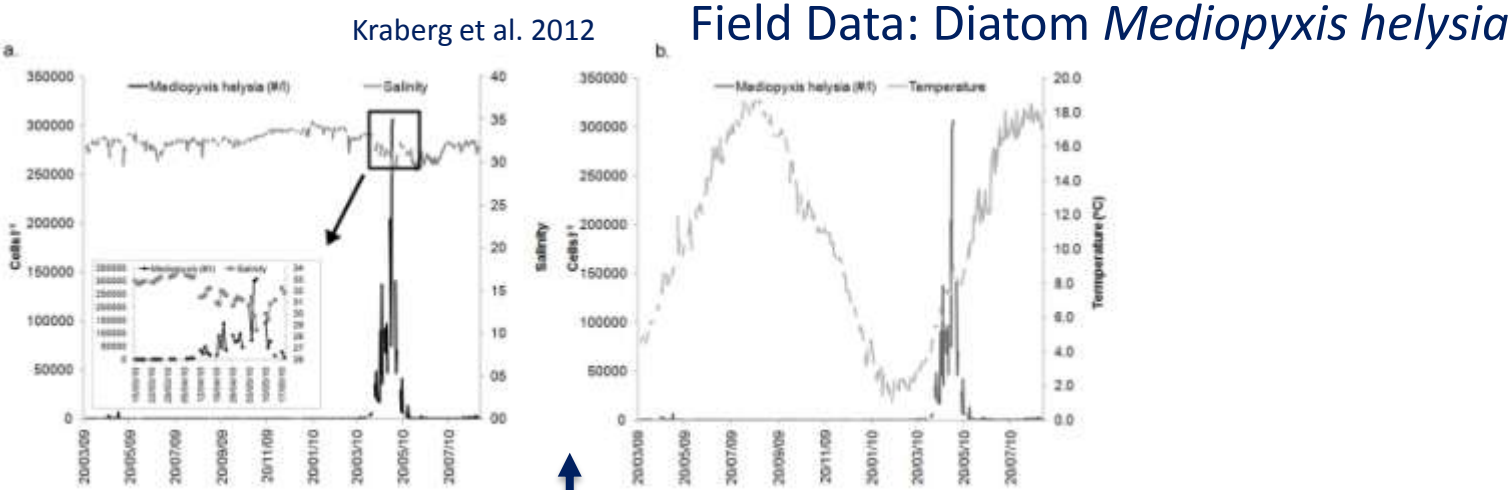
Odontella aurita (Lyngbye) C. A. Agardh, 1832

taxon details	taxon details
caption	Live cells of the centric diatom <i>Odontella aurita</i>
archive	download large image
image author	alexandra
image type	PHOTO
copyright	Creative Commons Attribution 3.0
observation details	click here
image notes	Live cells from net haul (20 µm mesh size), magnification x 400, Zeiss Axioskop 2
image sets	North Sea, Helgoland Roads Phytoplankton Monitoring
publish image	<input checked="" type="checkbox"/>
dataset	
responsible institution	Alfred Wegener Institute for Polar and Marine Research

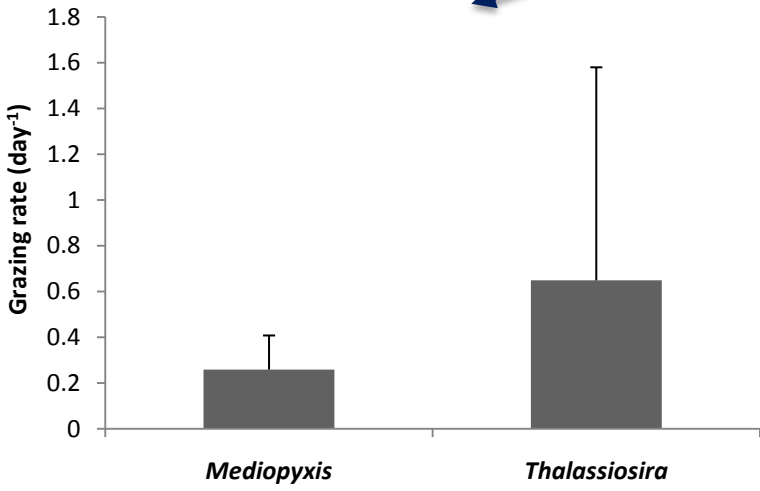
Linking taxonomic to numerical data



images

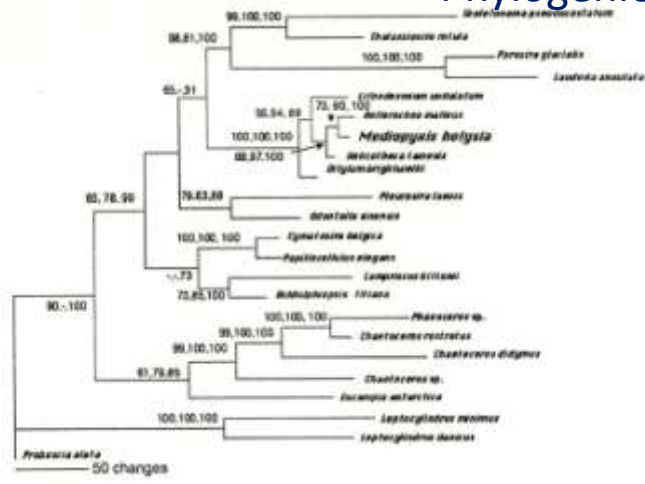


Ecological data



Contextual data

Phylogenies/molecular monitoring



From Medlin et al. 2006

Database systems: How can they help?



- They can provide a ‚taxonomic backbone‘
 - Can make data searchable by higher taxon
 - Provide taxonomic metadata for numerical datasets
 - Act as a reference for concepts used in a timeseries (reference species lists)
- Need to be able to maintain very complex metadata
 - Images and their metadata
 - Observational data with appropriate metadata standards
 - Analytical metadata (microscopic, photographic equipment)
 - Taxonomic/nomenclatural metadata

Taxonomic databases: Worms



The image shows two overlapping browser windows from the WoRMS website. The top window displays the 'WoRMS Taxon list' search results for 'odontella', listing various species like *Odontella C. Agardh, 1832* and *Odontella antediluviana*. The bottom window shows the 'WoRMS taxon details' for *Odontella aurita (Lyngbye) C. Agardh, 1832*. The details page includes classification (Diatia, Chlorista, etc.), status (accepted), rank (Species), parent (*Odontella C. Agardh, 1832*), synonymised taxa (*Diadubonia aurita*, *Diatoma aurita*, *Diatoma auritum*), sources (Muyllert, K.; Sabbe, K. 1999), basis of record (Algaebase), direct child taxa (*Odontella aurita* var. *minima*, *Odontella aurita* var. *obtusata*), environment (marine), and distribution (Belle des Chateaux, Bay of Fundy, FAO fishery area 21).

Taxonomic databases: Algaebase



[content](#) [about](#) [team](#) [links](#) [contact](#) [search](#)

[genus](#) · [species](#) · [literature](#) · [journals](#) · [images](#) · [common names](#) · [distribution](#) · [glossary](#) · [taxonomy browser](#) · [higher taxonomy](#)

134,256 species and infraspecific names are in the database, 16,826 images, 49,730 bibliographic items, 221,786 distributional records.

Thalassiosira punctigera (Castracane) Hasle

Publication details

Thalassiosira punctigera (Castracane) Hasle 1983: 602

Original publication: Hasle, G.R. (1983). *Thalassiosira punctigera* (Castr.) comb. nov., a widely distributed marine planktonic diatom. *Nordic Journal of Botany* 3: 593-608.

Type species

The type species (holotype) of the genus *Thalassiosira* is *Thalassiosira nordenskiöldii* Cleve.

Status of name

This name is of an entity that is currently accepted taxonomically.

Basionym

Ethmodiscus punctiger Castracane

Homotypic Synonym(s)

Ethmodiscus punctiger Castracane 1886

Coscinodiscus punctiger (Castracane) H.Peragallo 1889

Heterotypic Synonym(s)

Ethmodiscus japonicus Castracane

Coscinodiscus angstii Gran 1931

Thalassiosira japonica Kisselev 1935

Classification:

Empire Eukaryota
Kingdom Chromista
Phylum Heterokontophyta
Class Coscinodiscophyceae
Order Thalassiosirales
Family Thalassiosiraceae
Genus *Thalassiosira*

[Taxonomy](#)

[References](#)

[Submit Feedback](#)

[Submit Reference](#)

Links

[Genbank](#)

[Index Nominum Algarum](#)

[Google](#)

[Biodiversity Heritage Library](#)

Links to numerical data: PANGAEA/Planktonnet

PANGAEA
Publishing Network for Geosciences & Environmental Data

Data Description

Citation: Wiltshire, Karen Helen (2002): Abundance of dinoflagellates, dictyochophyceae, and prymnesiophyceae at time series station Helgoland Roads in 1994. Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Unpublished dataset #77552

Project(s): Biologische Anstalt Helgoland (BAH) %
Long-term Ecological Research at AWI (LTER) %

Coverage: West: 7.9000 ° East: 7.9000 ° South: 54.1883 ° North: 54.1883 °
Date/Time Start: 1994-01-01T07:58:00 ° Date/Time End: 1994-12-28T08:20:00 °
Minimum DEPTH_water: 0.5 m ° Maximum DEPTH_water: 0.5 m

Event(s): HelgolandRoads (Kabeltonne) % * Latitude: 54.1883 ° Longitude: 7.9000 ° Elevation: -10.0 m * Location: German Bight, North Sea %
* Campaign: HelgolandRoadsTimeseries % * Basis: Meeresstation Helgoland % * Device: Monitoring %

Further details: hdl:10013/epic:21653:dt01

Comment: Investigator of phytoplankton samples: Klaus Trutner

Parameter(s)

#	Name	Short Name	Unit	Principal Investigator	Method	Comment
1	DATE/TIME %	Date/Time				Seacode
2	DEPTH_water %	Depth/water	m			Seacode
3	Carotum fusca, fractionated %	C. fusca frac	#0	Wiltshire, Karen Helen %	Quantitative phytoplankton method (J.ferminh), 1950 %	size: 50-250 µm
4	Carotum fusus, fractionated %	C. fusus frac	#0	Wiltshire, Karen Helen %	Quantitative phytoplankton method (J.ferminh), 1950 %	size: 15-20 µm; hdl:10013/epic:21653:dt01:planktonnet:img:12595
5	Carotum hordium, fractionated %	C. hordium frac	#0	Wiltshire, Karen Helen %	Quantitative phytoplankton method (J.ferminh), 1950 %	size: 40-100 µm; hdl:10013/epic:21653:dt01:planktonnet:img:12597
6	Carotum lineatum, fractionated %	C. lineatum frac	#0	Wiltshire, Karen Helen %	Quantitative phytoplankton method (J.ferminh), 1950 %	size unknown; hdl:10013/epic:21653:dt01:planktonnet:img:14215
7	Carotum tripos, fractionated %	C. tripos frac	#0	Wiltshire, Karen Helen %	Quantitative phytoplankton method (J.ferminh), 1950 %	size: 80-250 µm; hdl:10013/epic:21653:dt01:planktonnet:img:12604
8	Dinophysis sp., fractionated %	Dinophysis sp. frac	#0	Wiltshire, Karen Helen %	Quantitative phytoplankton method (J.ferminh), 1950 %	class 1, size unknown
9	Dinophysis sp., fractionated %	Dinophysis sp. frac	#0	Wiltshire, Karen Helen %	Quantitative phytoplankton method (J.ferminh), 1950 %	class 2, size unknown
10	Dinophysis sp., fractionated %	Dinophysis sp. frac	#0	Wiltshire, Karen Helen %	Quantitative phytoplankton method (J.ferminh), 1950 %	class 3, size unknown
11	Dinophysis sp. %	Dinophysis sp.	#0	Wiltshire, Karen Helen %	Quantitative phytoplankton method (J.ferminh), 1950 %	
12	Gonodinium sp., fractionated %	Gonodinium sp. frac	#0	Wiltshire, Karen Helen %	Quantitative phytoplankton method (J.ferminh), 1950 %	class 1, size unknown
13	Gonodinium sp., fractionated %	Gonodinium sp. frac	#0	Wiltshire, Karen Helen %	Quantitative phytoplankton method (J.ferminh), 1950 %	class 2, size unknown
14	Gonodinium sp., fractionated %	Gonodinium sp. frac	#0	Wiltshire, Karen Helen %	Quantitative phytoplankton method (J.ferminh), 1950 %	class 4, size unknown
15	Gonodinium sp., fractionated %	Gonodinium sp. frac	#0	Wiltshire, Karen Helen %	Quantitative phytoplankton method (J.ferminh), 1950 %	class 7, size unknown

<http://pangaea.de>

- PANGAEA: Each IDENTIFIED taxon labelled with a representative image in PLANKTON*NET
- Embedded in taxonomic hierarchy
- Images need standardization
 - Retrospect ,evaluation‘ of a record



<http://planktonnet.awi.de> 

Databases: Where do we stand?



- **Pangaea (<http://pangaea.de>)** : Excellent archival tool but no taxonomic backbone
- **EoL (<http://eol.org>)** : Excellent taxonomic resource but not specializing in geo-referenced data (but what is there is image-based, no link to numeric data)
- **Marine species/Eurobis (<http://marinespecies.org>)**
- **Algaebase (<http://algaebase.org>)**. A comprehensive taxonomic resource with, descriptions, references and images
- **PLANKTON*NET (<http://planktonnet.awi.de>)**: Geo-referenced, image based presence-absence data on plankton diversity

Summary: The ideal database



- Safe, persistent archival including all metadata
- Taxonomic backbone, contextual information,
- Images as metadata (e.g. PLANKTON*NET)
- Facilitate quality control
 - Facilitate comparisons between datasets
- Data products for visualization
 - Requires georeferenced information