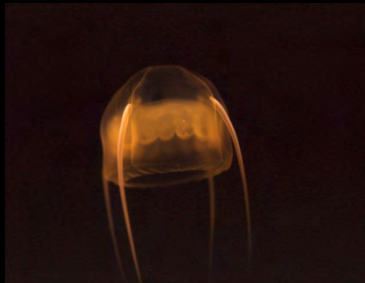




# The perils of bad taxonomy for leading edge science: a case study with the genus *Aegina*, and the consequences for Deep Learning



Dhugal Lindsay (JAMSTEC),

Mary Grossmann, Mitsuko Hidaka (JAMSTEC/Kitasato U.), Jun Nishikawa (Tokai U.), Hiroshi Miyake (Kitasato U.), Ryo Minemizu, Russell Hopcroft (U. Alaska), Bastian Bentlage (U. Guam), Allen Collins (Smithsonian), Takehisa Yamakita (JAMSTEC), Hiroyuki Yamamoto (JAMSTEC)

# *Aegina citrea* – Case Study

Treated as a single species for the last >50 years



Narcomedusae  
Aeginidae

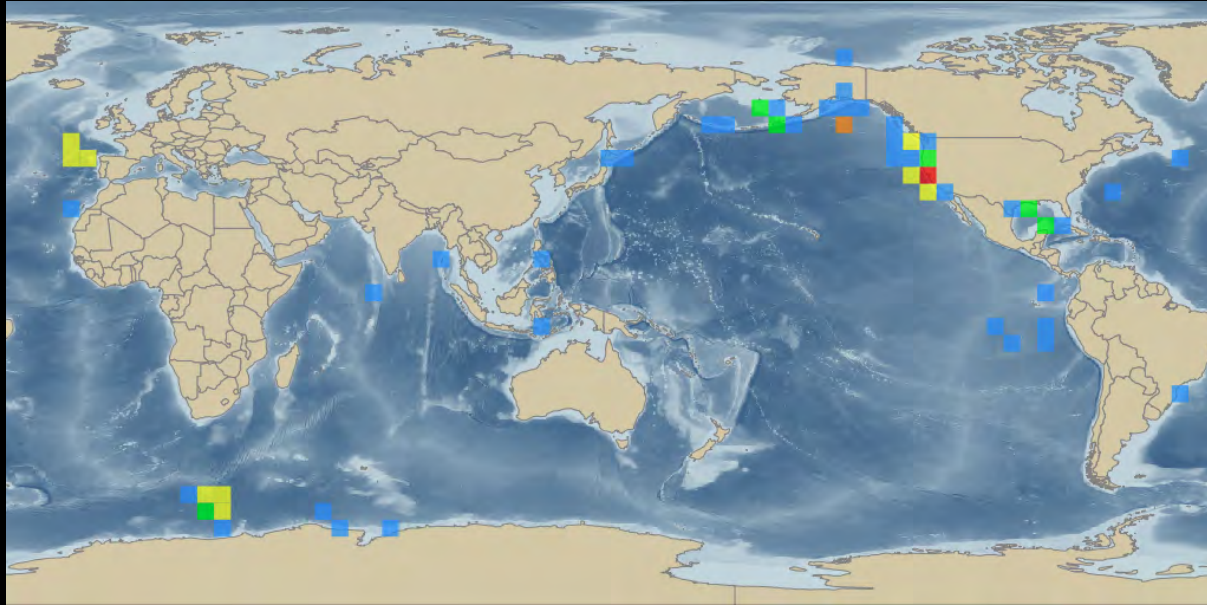
Kramp (1961) synonymized

- 1 *Aegina rosea* Eschscholtz, 1829
- 2 *Aegina rhodina* Haeckel, 1879
- 3 *Aegina eschscholtzii* Haeckel, 1879
- 4 *Aegina lactea* Vanhöffen, 1908
- 5 *Aegina brunnea* Vanhöffen, 1908
- 6 *Aegina alternans* Bigelow, 1909
- 7 *Aegina pentanema* Kishinouye, 1910
- 8 *Cunarcha aeginoides* Haeckel, 1879

Basically, any narcomedusa with 4-6 tentacles, 8-12 stomach pouches and a peripheral canal system were all relegated to this one species



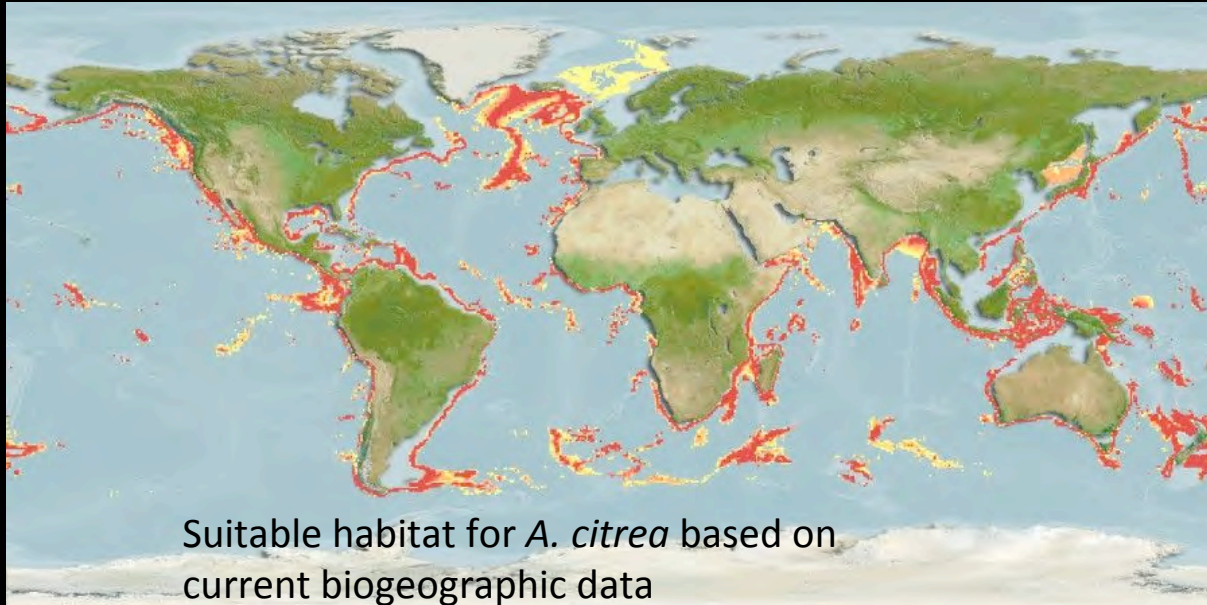
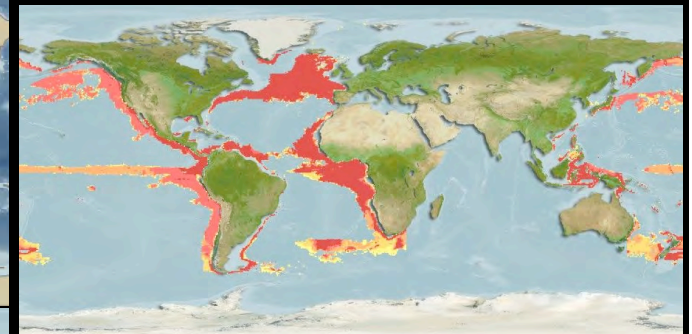
*Aegina citrea*: A Poster Child  
for determining environmental niches and  
assessing the effects of climate change



*Aegina citrea*

What data is in OBIS?

*Aegina citrea* (Year 2050 range)

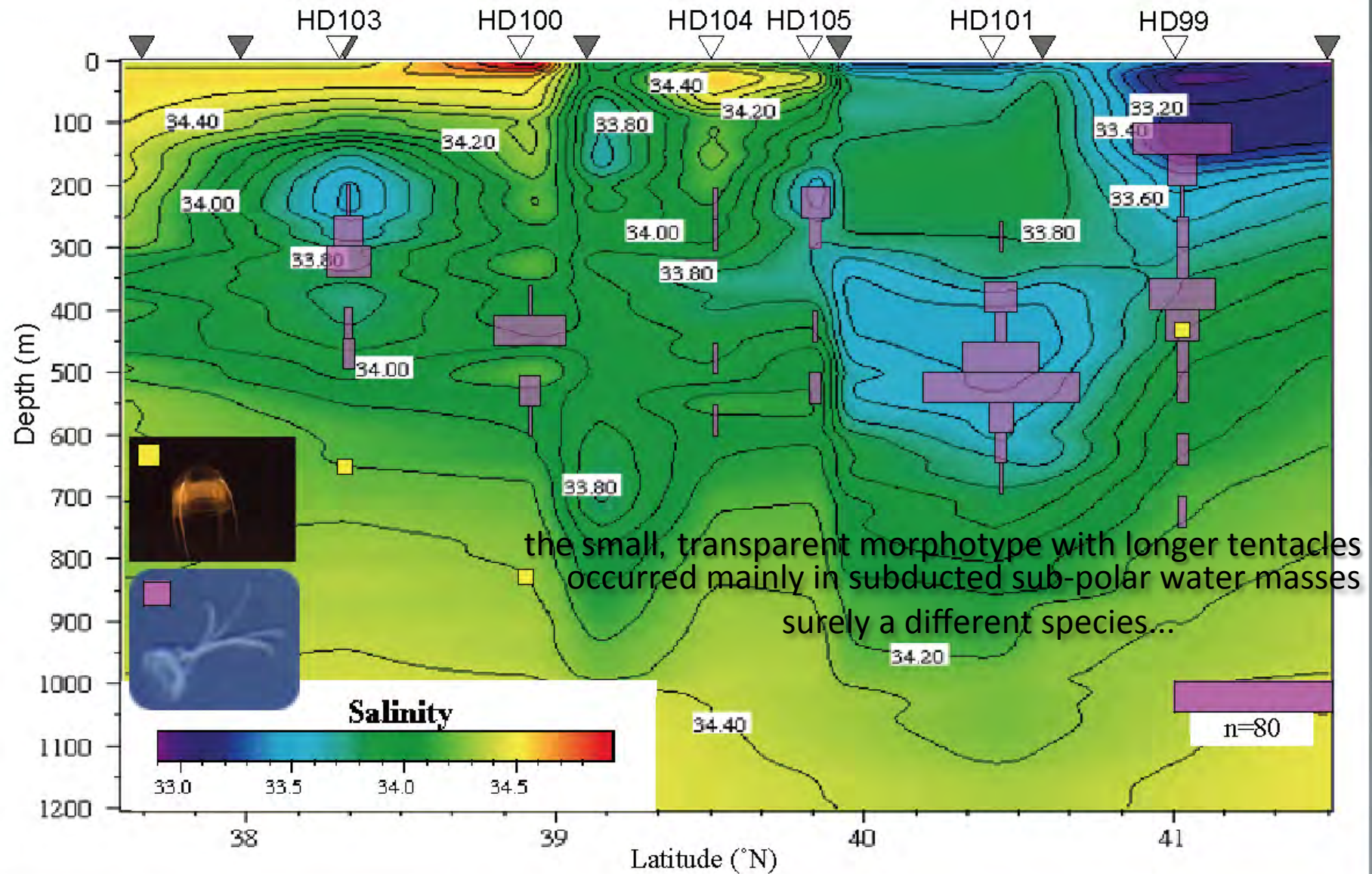


Suitable habitat for *A. citrea* based on  
current biogeographic data





# Only with Correct Taxonomy can Niches be characterised



Horizontal and vertical distribution of "*Aegina citrea*" over the Japan Trench from 20 April to 6 May 2002. Salinity profiles obtained using a Seabird SBE-19 CTD mounted on the ROV HYPER-DOLPHIN (white triangles) and X-CTD probes deployed between dive points (grey triangles).

# Phylogenetics of Trachylina (Cnidaria: Hydrozoa) with new insights on the evolution of some problematical taxa

ALLEN G. COLLINS<sup>1</sup>, BASTIAN BENTLAGE<sup>2</sup>, ALBERTO LINDNER<sup>3</sup>, DHUGAL LINDSAY<sup>4</sup>, STEVEN H.D. HADDOCK<sup>5</sup>, GERHARD JARMS<sup>6</sup>, JON L. NORENBURG<sup>7</sup>, THOMAS JANKOWSKI<sup>8</sup> AND PAULYN CARTWRIGHT<sup>2</sup>

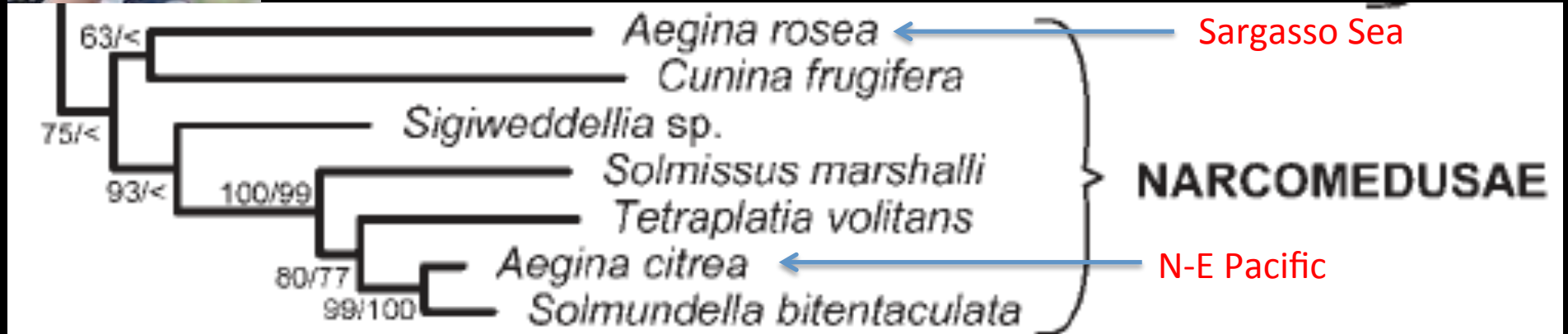
We pointed out in this 2008 paper that *Aegina citrea* and *A. rosea* were genetically different enough to be put in 2 separate families



our "A. rosea"



our "A. citrea"







# However to our dismay...

Aegina rosea small subunit 18S ribosomal RNA gene, partial sequence

GenBank: EU247813.1

FASTA Graphics PopSet

Go to:

LOCUS EU247813 1739 bp DNA linear INV 05-MAY-2014

DEFINITION Aegina rosea small subunit 18S ribosomal RNA gene, partial sequence.

ACCESSION EU247813

VERSION EU247813.1 GI:166007992

KEYWORDS .

SOURCE Aegina citrea

ORGANISM Aegina citrea

Eukaryota; Metazoa; Cnidaria; Hydrozoa; Trachylina; Narcomedusae; Aeginidae; Aegina.

REFERENCE 1 (bases 1 to 1739)

AUTHORS Collins,A.G., Bentlage,B., Lindsay,D., Haddock,S.H.D., Lindner,A., Norenburg,J.L., Jarms,G., Jankowski,T. and Cartwright,P.

TITLE Phylogenetics of Trachylina (Cnidaria: Hydrozoa)

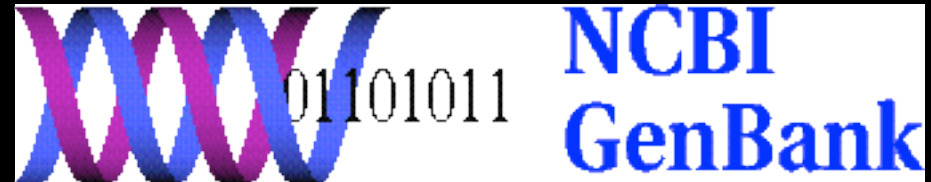
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1739)

AUTHORS Collins,A.G., Bentlage,B., Lindsay,D., Haddock,S.H.D., Lindner,A., Norenburg,J.L., Jarms,G., Jankowski,T. and Cartwright,P.

TITLE Direct Submission

JOURNAL Submitted (25-OCT-2007) Dept. of Ecology and Evolutionary Biology, The University of Kansas, 1200 Sunnyside Avenue, Lawrence, KS 66045, USA



*Aegina rosea* small subunit 18S ribosomal RNA gene



somehow our *Aegina rosea* sequence has been labelled *Aegina citrea*, even though our paper described family level differences and we registered it as *Aegina rosea*!!!

FEATURES	Location/Qualifiers
source	1..1739
	/organism="Aegina citrea"
	/mol_type="genomic DNA"
	/specimen_voucher="JAMSTEC; dive, RB-MOC1-001"
	/db_xref="taxon:168710"
rRNA	<1..>1739
	/product="18S small subunit ribosomal RNA"



# ITIS

## Integrated Taxonomic Information System

Two main taxonomic databases that GenBank refers to for its own taxonomic database

Home

About ITIS

Welcome to ITIS, the Integrated Taxonomic Information System! Here you will find authoritative taxonomic information on plants, animals, fungi, and microbes of North America and the world. We are a [partnership](#) of U.S., [Canadian](#), and [Mexican](#) agencies ([ITIS-North America](#)); other organizations; and taxonomic specialists. ITIS is also a partner of [Species 2000](#) and the [Global Biodiversity Information Facility \(GBIF\)](#). The ITIS and Species 2000 [Catalogue of Life \(CoL\)](#) partnership is proud to provide the taxonomic backbone to the [Encyclopedia of Life \(EOL\)](#).



# WoRMS

World Register of Marine Species

Home	<b>WoRMS taxon details</b>
About	
Search taxa	✓ <b><i>Aegina citrea</i> Eschscholtz, 1829</b>
Taxon tree	AphiaID: 117263
Literature	<b>Classification:</b> Biota > ✓ Animalia (Kingdom) > ✓ Cnidaria (Phylum) > ✓ Hydrozoa (Class) > ✓ Trachylinae (Subclass) > ✓ Narcomedusae (Order) > ✓ Aeginidae (Family) > ✓ <i>Aegina</i> (Genus)
Distribution	
Specimens	<b>Status</b> accepted
Match taxa	<b>Rank</b> Species
Editors	<b>Parent</b> ✓ <i>Aegina</i> Eschscholtz, 1829
Statistics	<b>Synonymised names</b> ✓ <i>Aegina rosea</i> Eschscholtz, 1829 (Synonym)
Users	
Webservice	<b>Sources</b> <b>original description</b> Eschscholtz, F. 1829. System der Acalephen. Eine ausführliche Beschreibung aller medusenartigen Strahltiere. Ferdinand Dümmler, Berlin, pp. 1-190, 116 pls., available online at <a href="http://www.biodiversitylibrary.org/item/40493#page/5/mode/1up">http://www.biodiversitylibrary.org/item/40493#page/5/mode/1up</a> page(s): 113 [details]
Photogallery	
Info downloads	[show all]
Sponsors	<b>Vernacular Names</b>
Glossary	<b>Language</b> <b>Name</b> English ? golf tee medusa [details]
Manual	<b>Environment</b> marine



[Go to Print Version](#)

***Aegina* Eschscholtz, 1829** in ITIS *Aegina* contains 3 valid species

Taxonomic Serial No.: 51145

Download data: [\(Download Help\)](#) *Aegina* TSN 51145

**Taxonomy and Nomenclature**

Kingdom: Animalia  
 Taxonomic Rank: Genus  
 Synonym(s):  
 Common Name(s):

**Taxonomic Status:**

Current Standing: valid

**Data Quality Indicators:**

Record Credibility Rating: verified - standards met  
 Global Species Completeness: partial  
 Latest Record Review: 2006

**Taxonomic Hierarchy**

Kingdom	<a href="#">Animalia</a> – Animal, animaux, animals
Subkingdom	<a href="#">Radiata</a>
Phylum	<a href="#">Cnidaria</a> Hatschek, 1888 – cnidarians, coelenterates, cnidaires, coelentérés, água viva, anêmona, caravela, cnidario, coral, hidra
Subphylum	<a href="#">Medusozoa</a>
Class	<a href="#">Hydrozoa</a> Owen, 1843 – hydralike animals, hydroids, hydrozoans, hydraires, hydrozoaires, água viva, hidra, hidrozoa, hidrozoário, pólipo
Subclass	<a href="#">Trachylina</a>
Order	<a href="#">Narcomedusae</a> Haeckel, 1879
Suborder	<a href="#">Narcomedusida</a>
Family	<a href="#">Aeginidae</a>
Genus	<i>Aegina</i> Eschscholtz, 1829
	<b>Direct Children:</b>
Species	<a href="#">Aegina citrea</a> Eschscholtz, 1829 – golf tee medusa
Species	<a href="#">Aegina longicornis</a>
Species	<a href="#">Aegina rosea</a>

ITIS taxonomy is based on the NODC Taxonomic Code



**NOAA** NATIONAL OCEANOGRAPHIC DATA CENTER (NODC)  
 UNITED STATES DEPARTMENT OF COMMERCE

last updated on 13 June 1996 (and not reviewed since...)

**Other Source(s):**

Source: [NODC Taxonomic Code, database \(version 8.0\)](#)  
 Acquired: 1996

3 valid species!

*Aegina citrea* from type locality (off Japan)



"*Aegina rosea*" of Collins et al 2008 following Bigelow, 1913



- Home
- About
- Search taxa
- Taxon tree
- Literature
- Distribution
- Specimens
- Match taxa
- Editors
- Statistics
- Users
- Webservice
- Photogallery
- Info downloads
- Sponsors
- Glossary
- Manual
- Log in

---

- RSS
- Add provider
- @WRMarineSpecies

## WoRMS taxon details

### ▼ *Aegina citrea* Eschscholtz, 1829

AphiaID: 117263

**Classification:** Biota > [Animalia](#) (Kingdom) > [Cnidaria](#) (Phylum) > [Hydrozoa](#) (Class) > [Trachylinae](#) (Subclass) > [Narcomedusae](#) (Order) > [Aeginidae](#) (Family) > [Aegina](#) (Genus)

**Status:** accepted

**Rank:** Species

**Parent:** [Aegina Eschscholtz, 1829](#)

**Synonymised names:** [Aegina rosea Eschscholtz, 1829](#) (Synonym)

**Sources:** **original description** Eschscholtz, F. 1829. System der Acalephen. Eine ausführliche Beschreibung aller medusenartigen Strahltiere. Ferdinand Dümmler, Berlin, pp. 1-190, 116 pls., available online at <http://www.biodiversitylibrary.org/item/40493#page/5/mode/1up> page(s): 113 [details]

[show all]

**Vernacular Names:**

Language	Name
English	? golf tee medusa [details]

**Environment:** marine

**Distribution:** **FROM OTHER SOURCES**

**Gulf of Mexico**

- ? Gulf of Mexico [details]

**Mediterranean Sea**

- ? Mediterranean Sea [details]

**North Atlantic Ocean**

- ? European waters (ERMS scope) [details]
- ? North West Atlantic [details]
- ? United Kingdom Exclusive Economic Zone [details]

**North Pacific Ocean**

- ? North Pacific [details]

**(no group)**

- ? Canadian Exclusive Economic Zone [Pacific part] [details]
- ? FAO fishing area 67 [details]
- ? New Zealand Exclusive Economic Zone [details]
- ? North East Pacific [details]


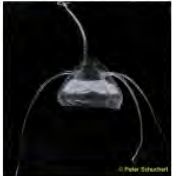
**Links:**

- ? To Biodiversity Heritage Library (23 publications)
- ? To Encyclopedia of Life
- ? To GenBank (4 nucleotides; 0 proteins)
- ? To Marine Species Identification Portal
- ? To PEST
- ? To USNM Invertebrate Zoology Cnidaria Collection
- ? To ITIS

**Notes:** **FROM OTHER SOURCES**

- ? **Biology** direct development [details]
- ? **Distribution** cosmopolitan [details]

**Images:**

[Aegina citrea](#)

[Aegina citrea, Puget S...](#)

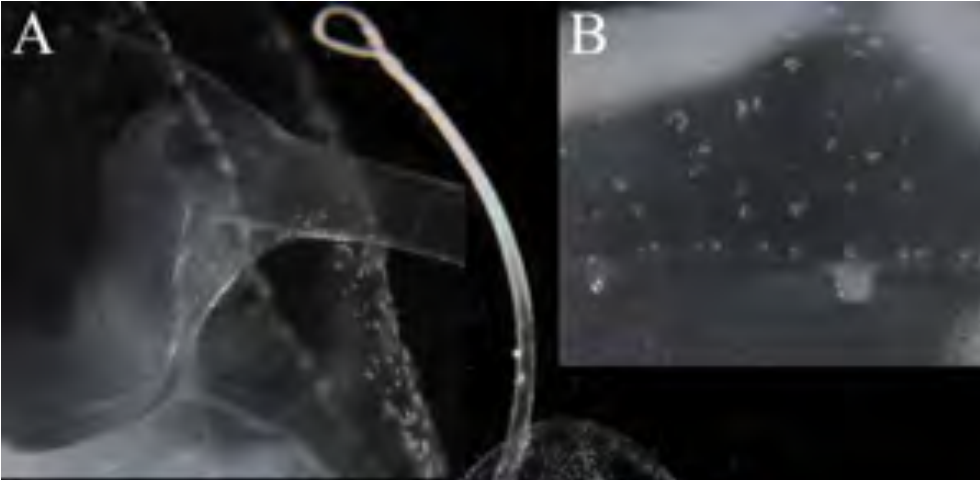
*A. rosea* treated as a synonym of *A. citrea*

GenBank must follow this





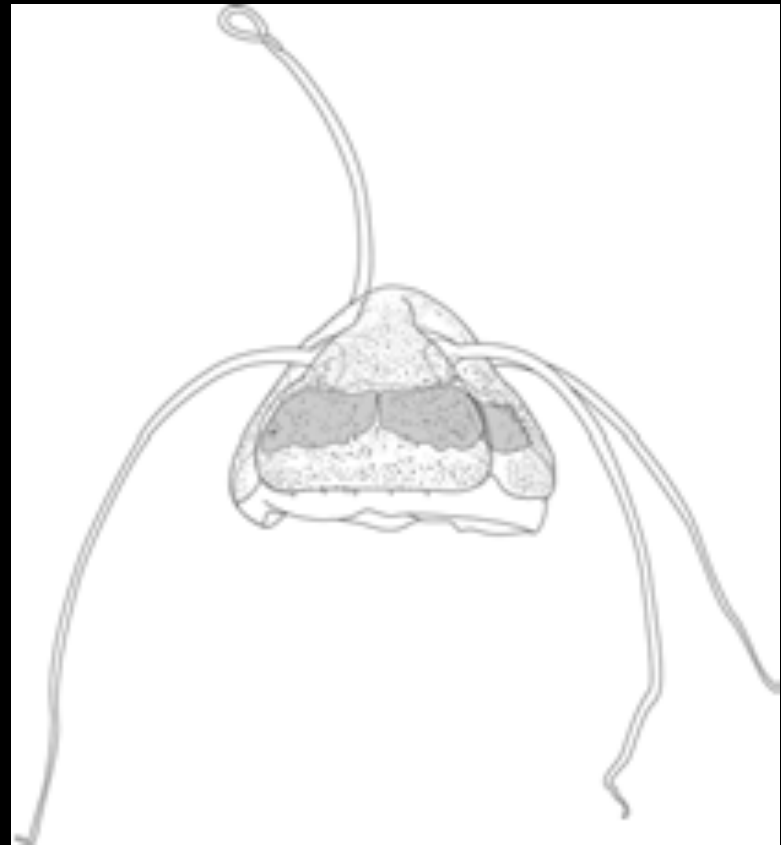
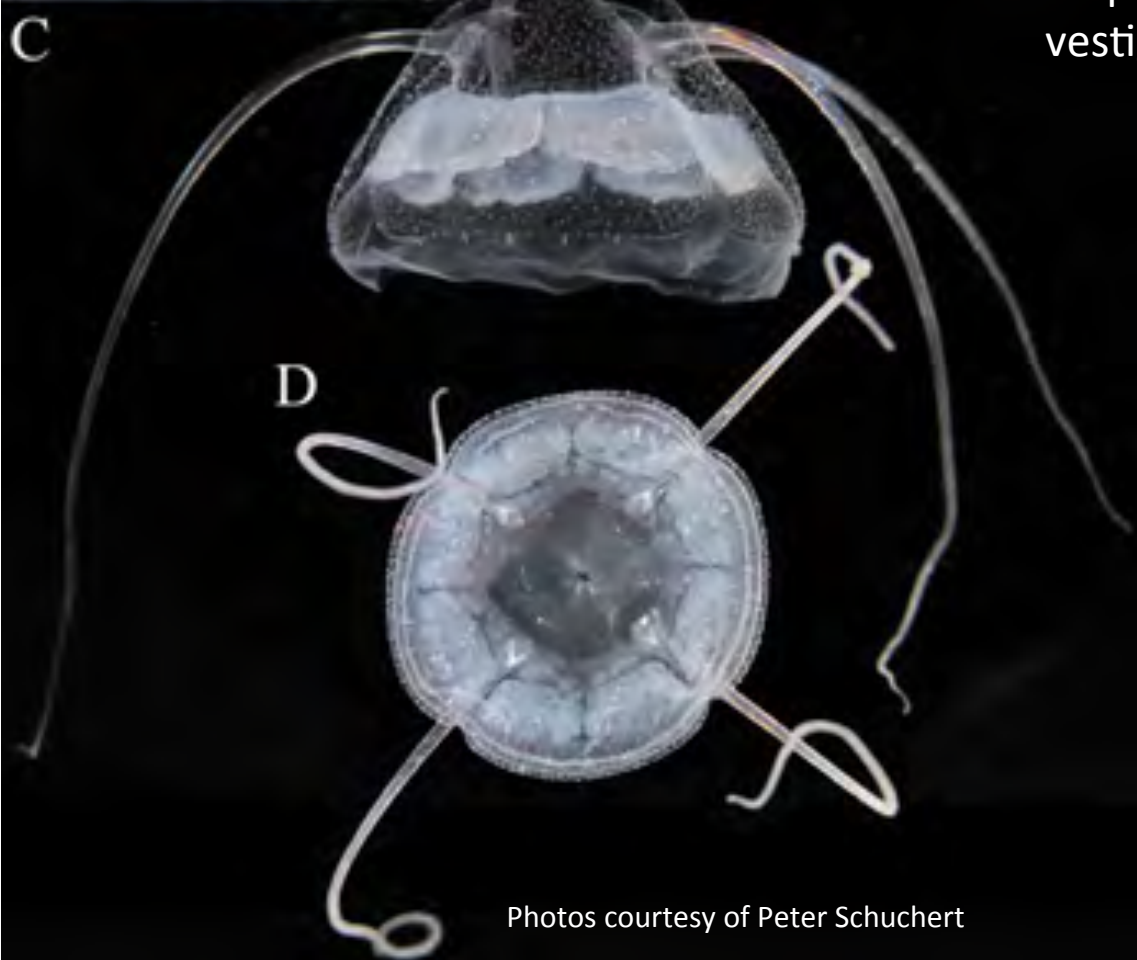




Solmundaeginidae Lindsay, Bentlage & Collins, 2017  
*Solmundaegina* Lindsay, 2017  
*Solmundaegina nematophora* Lindsay, 2017

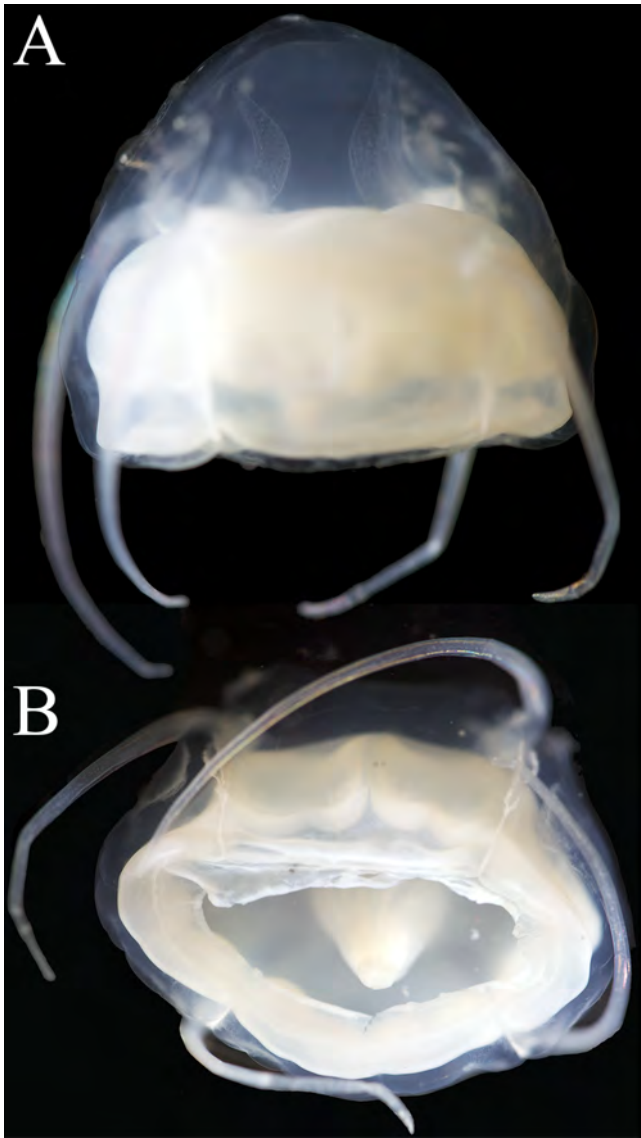
**NEW SPECIES, GENUS & FAMILY!!**

small downward-pointing tentacle roots  
no peripheral canal system  
vestigial secondary tentacle bulbs



Photos courtesy of Peter Schuchert

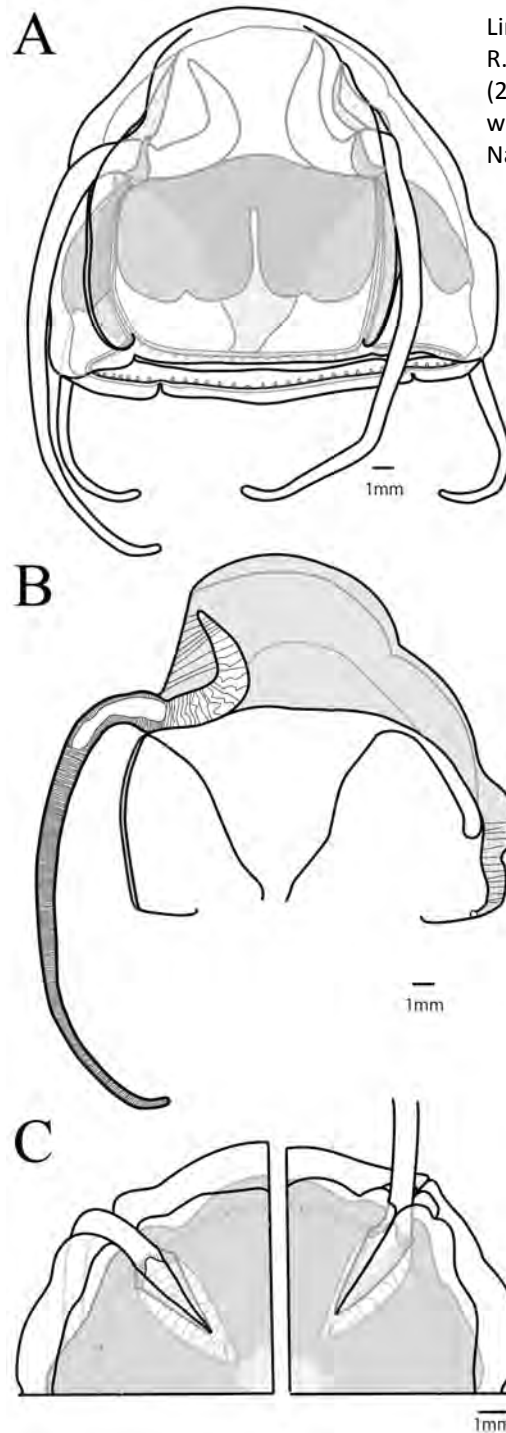




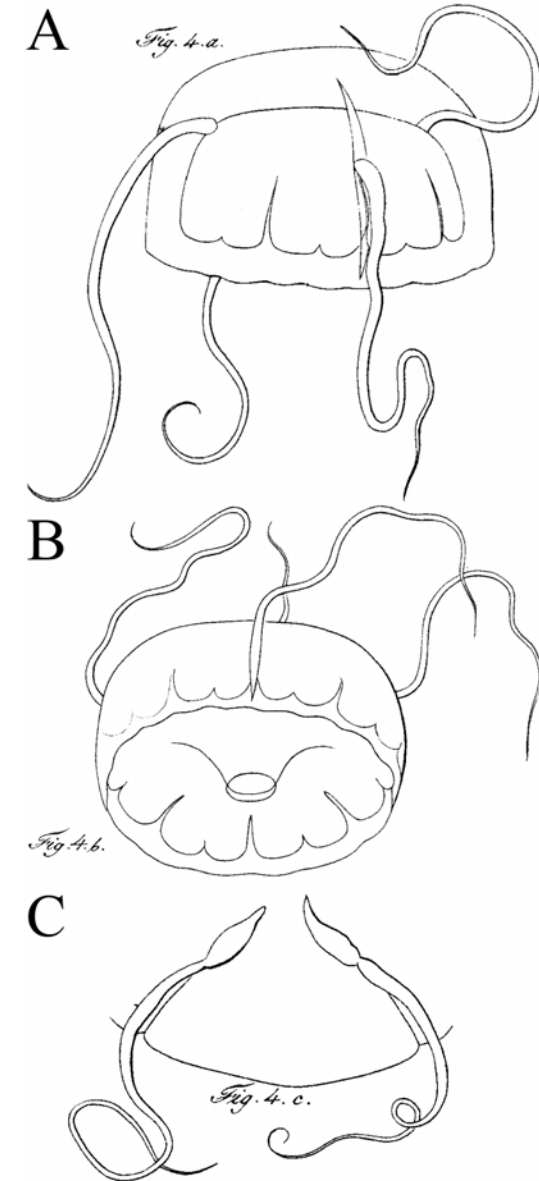
## The REAL *Aegina citrea*!

from near the Type Locality off eastern Japan

- large upward-pointing tentacle roots
- peripheral canal system
- no vestigial secondary tentacle bulbs
- deep apical grooves



Lindsay, D.J., Grossmann, M.M., Bentlage, B., Collins, A.G., Minemizu, R., Hopcroft, R.R., Miyake, H., Hidaka-Umetsu, M. and Nishikawa, J. (2017) The perils of online biogeographic databases: A case study with the "monospecific" genus *Aegina* (Cnidaria, Hydrozoa, Narcomedusae). *Marine Biology Research* 13(5): 494-512.



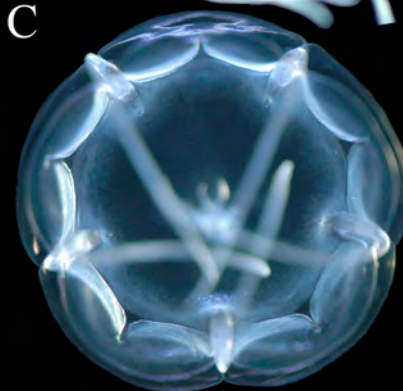
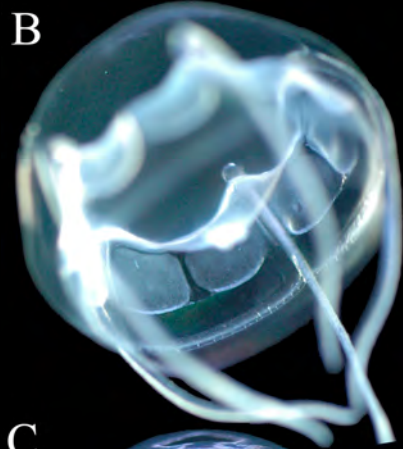




Pseudaeginidae Lindsay, Bentlage & Collins, 2017

*Pseudaegina* Lindsay, 2017

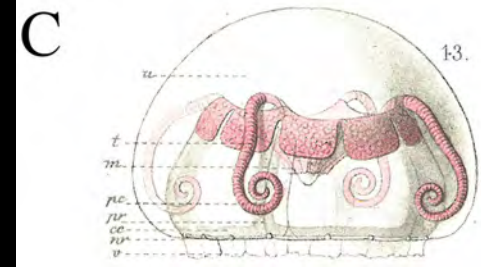
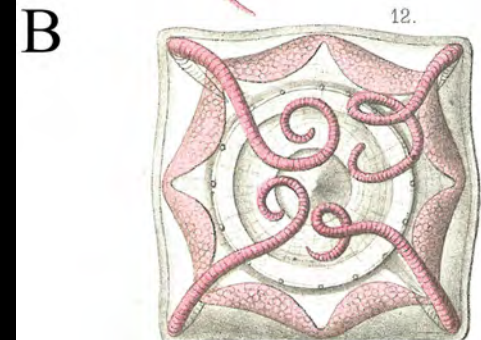
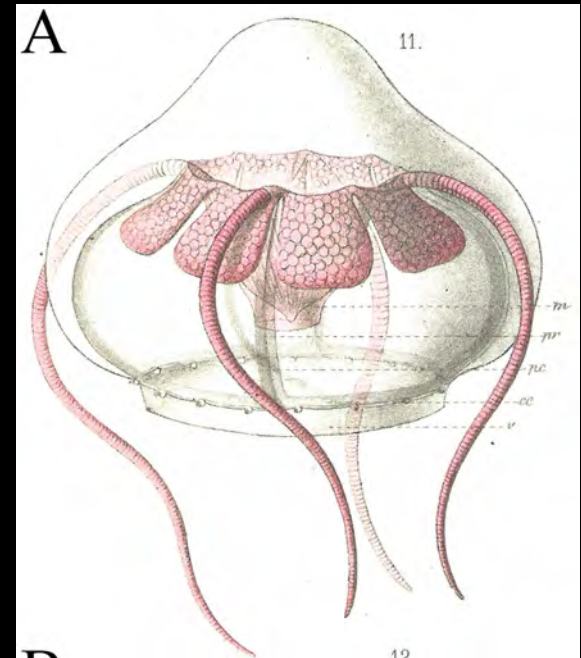
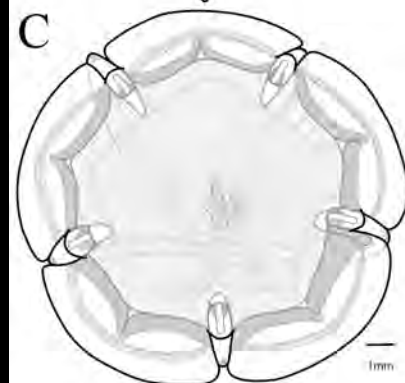
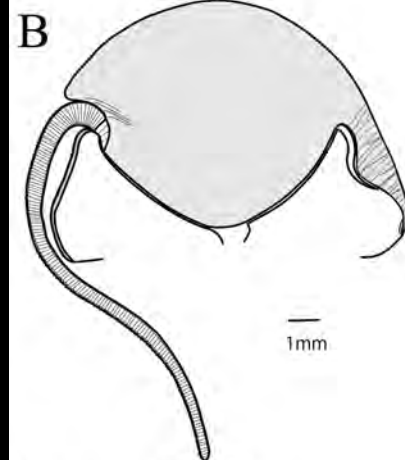
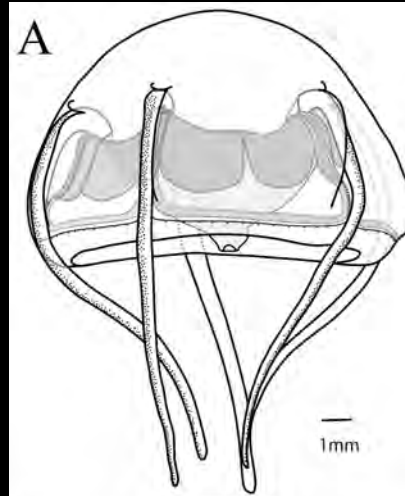
*Pseudaegina rhodina* (Haeckel, 1879)

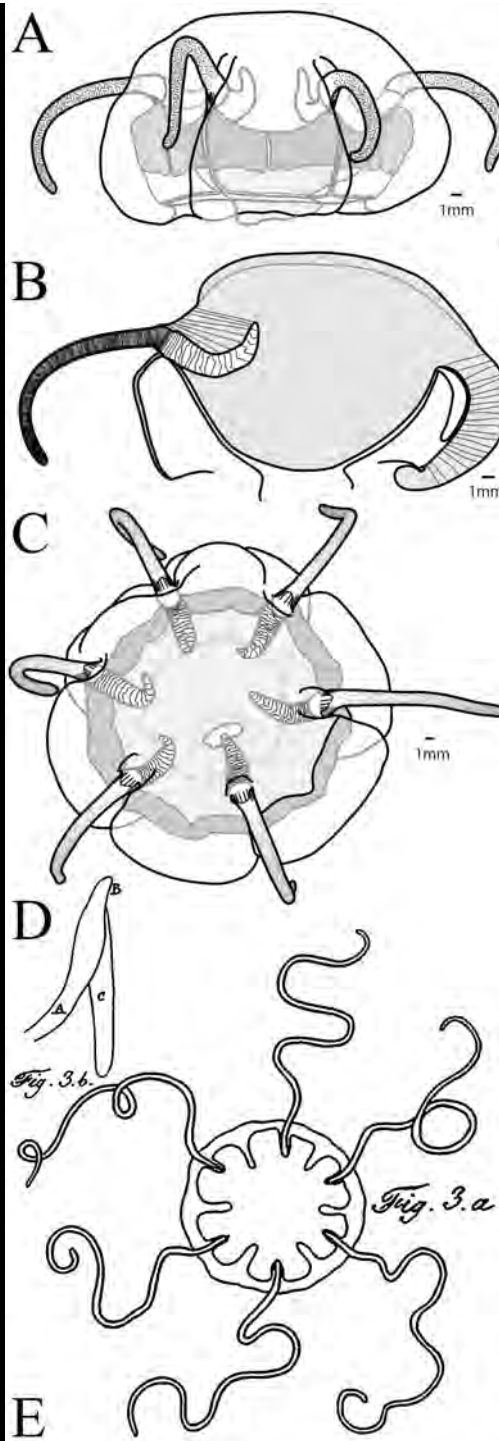
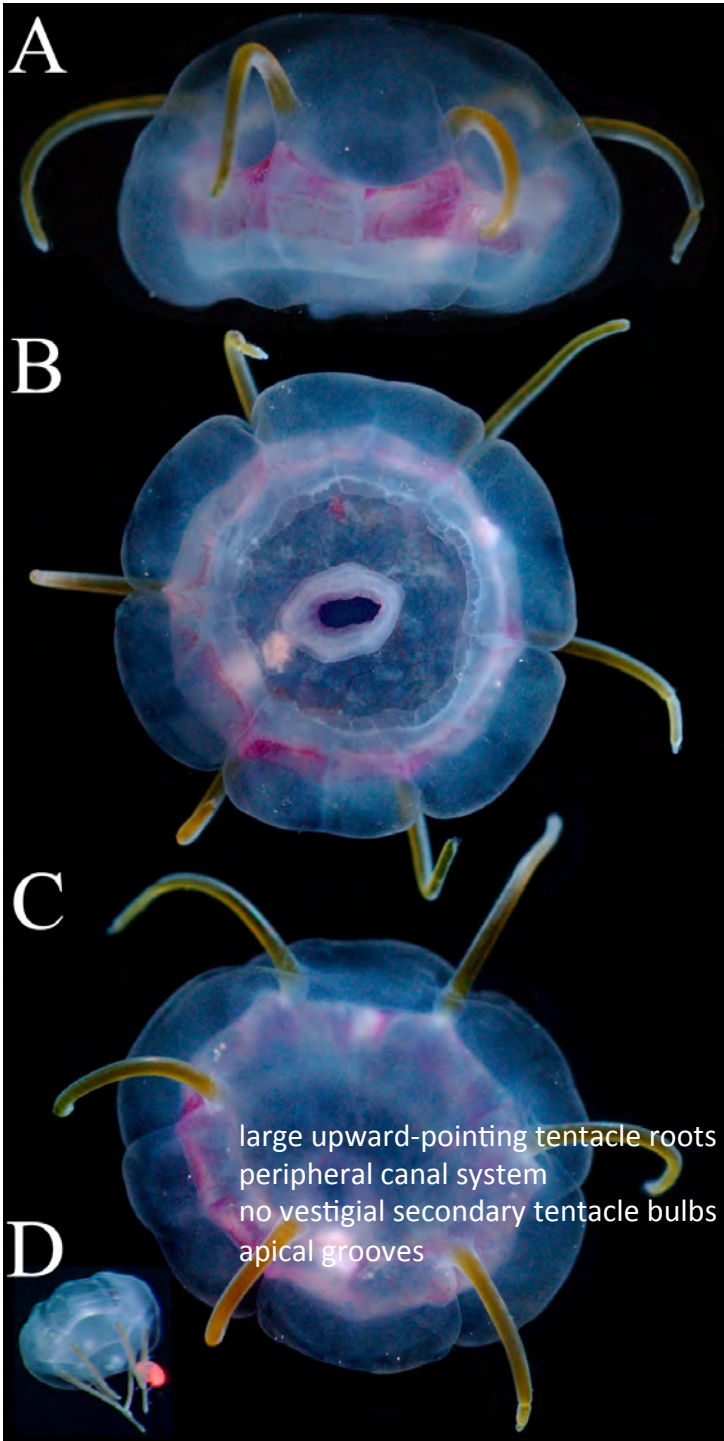


**NEW GENUS  
& FAMILY!!**

One of the *Aegina citrea* sequences in GenBank (the *A. rosea* of Collins et al 2008) is actually this species.

small downward-pointing tentacle roots  
no apical grooves  
peripheral canal system  
nematocysts only on upper surface of tentacles  
orally bulging mesoglea





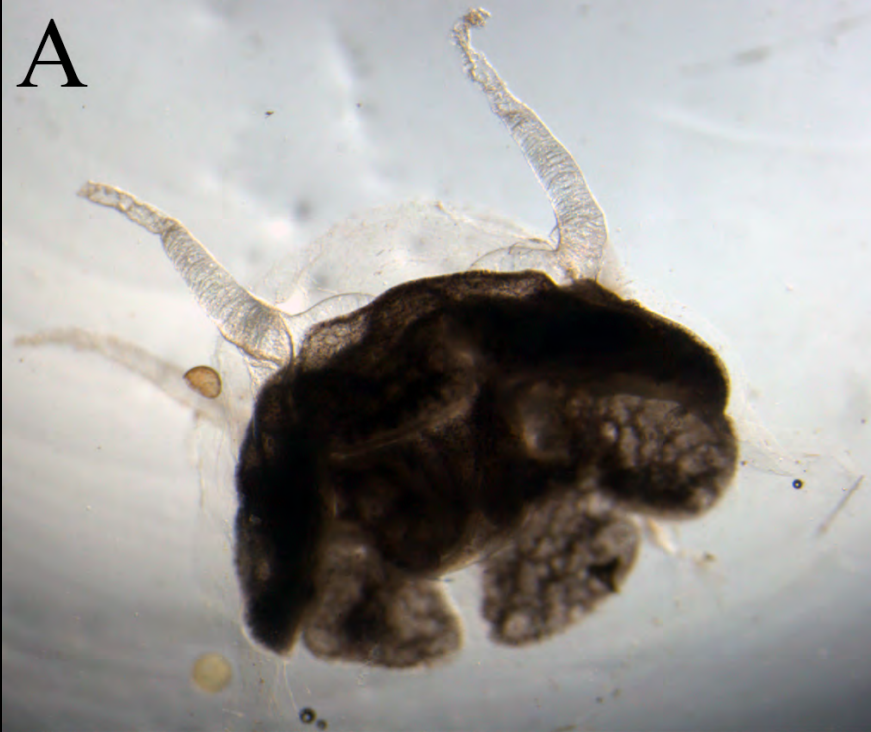
The REAL *Aegina rosea*!



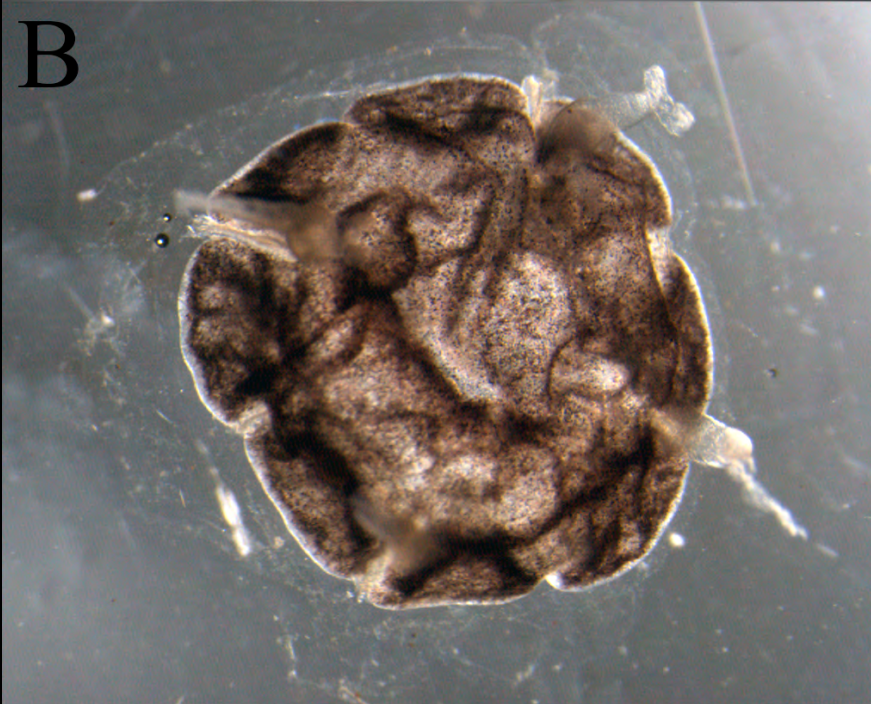
Therefore neither of the Collins et al 2008 sequences of (according to GenBank) "*Aegina citrea*" are actually that species. Furthermore, the species we labelled *A. rosea* actually belongs to a different family...



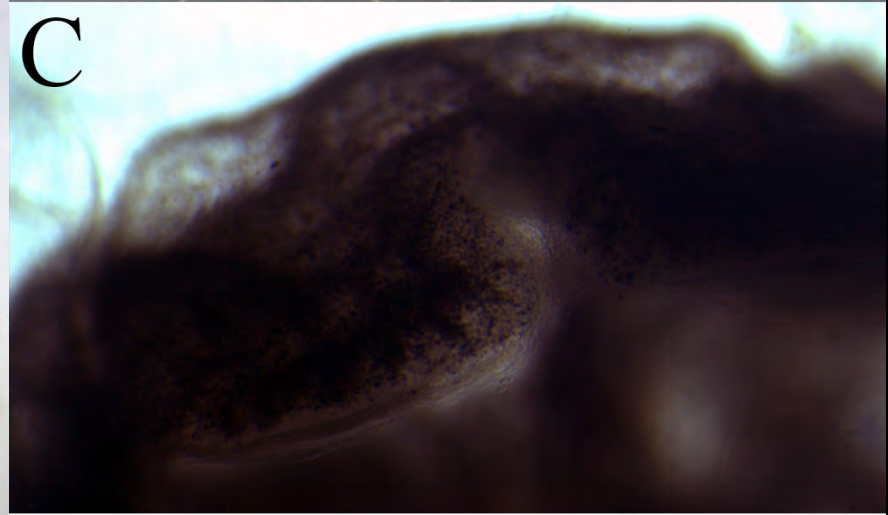
A



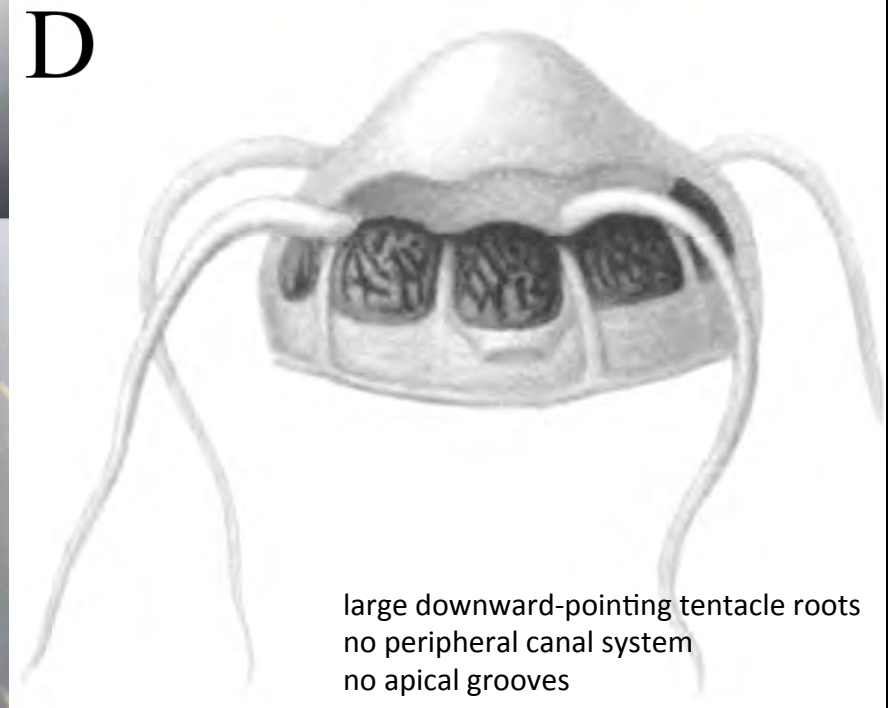
B



C

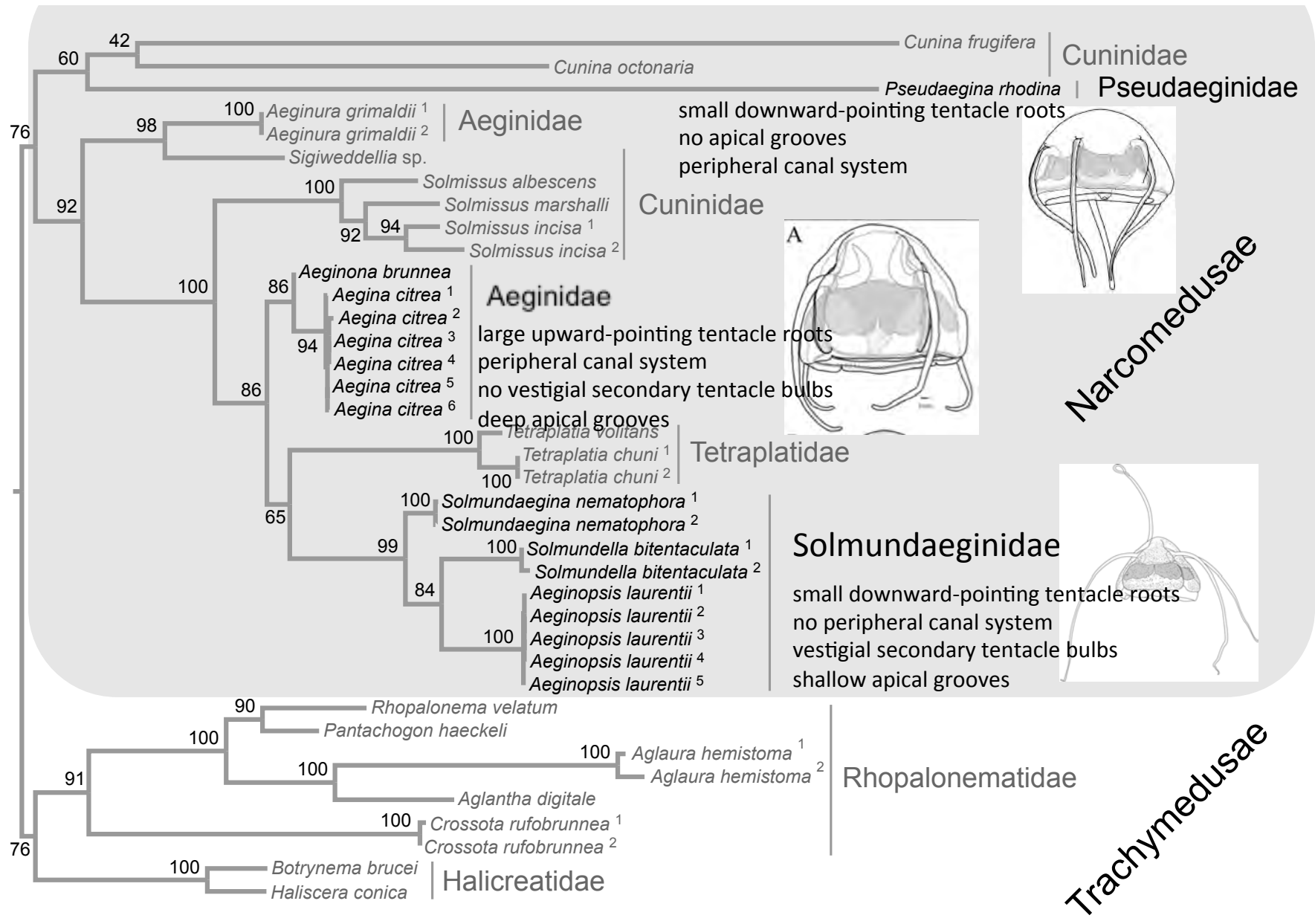


D



*Aeginona* Lindsay, 2017

*Aeginona brunnea* (Vanhöffen, 1908)



overlooked morphological characters actually most important!

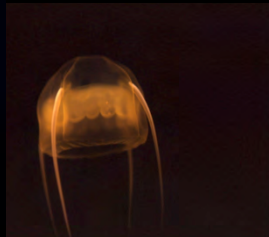
(ML topology based on combined 18S, 16S and COI data)

Lindsay, D.J., Grossmann, M.M., Bentlage, B., Collins, A.G., Minemizu, R., Hopcroft, R.R., Miyake, H., Hidaka-Umetsu, M. and Nishikawa, J. (2017) The perils of online biogeographic databases: A case study with the “monospecific” genus *Aegina* (Cnidaria, Hydrozoa, Narcomedusae). *Marine Biology Research* 13(5): 494-512.



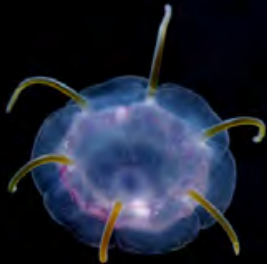
# Summary (Part 1)

- *Aegina citrea* was split into 6 separate species



# Summary (Part 1)

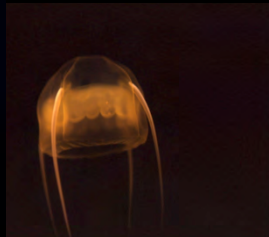
- *Aegina citrea* was split into 6 separate species
- Two new families, 3 new genera and 1 new species was described





# Summary (Part 1)

- *Aegina citrea* was split into 6 separate species
- Two new families, 3 new genera and 1 new species was described
- All *Aegina citrea* sequences in GenBank actually belonged to different species in different families to the real *Aegina citrea*



# Summary (Part 1)

- *Aegina citrea* was split into 6 separate species
- Two new families, 3 new genera and 1 new species was described
- All *Aegina citrea* sequences in GenBank actually belonged to different species in different families to the real *Aegina citrea*
- >90% of records in biogeographic databases are wrong





# Summary (Part 1)

- The data compromised by not having good taxonomy and systematics was not only biogeographic in nature, but also led to erroneous data on predator–prey interactions (e.g. Mills & Miller 1984, Choy et al. 2017), parasite–host associations (e.g. Gasca et al. 2007), behaviour and life history strategies (e.g. Larson et al. 1989), vertical distributions (e.g. Arai & Mason 1982), environmental factors driving distributions (e.g. Luo et al. 2014) and DNA barcode sequences (e.g. Collins et al. 2008). This case clearly illustrates how systematics and taxonomy provide the foundation upon which all other biological science is built.

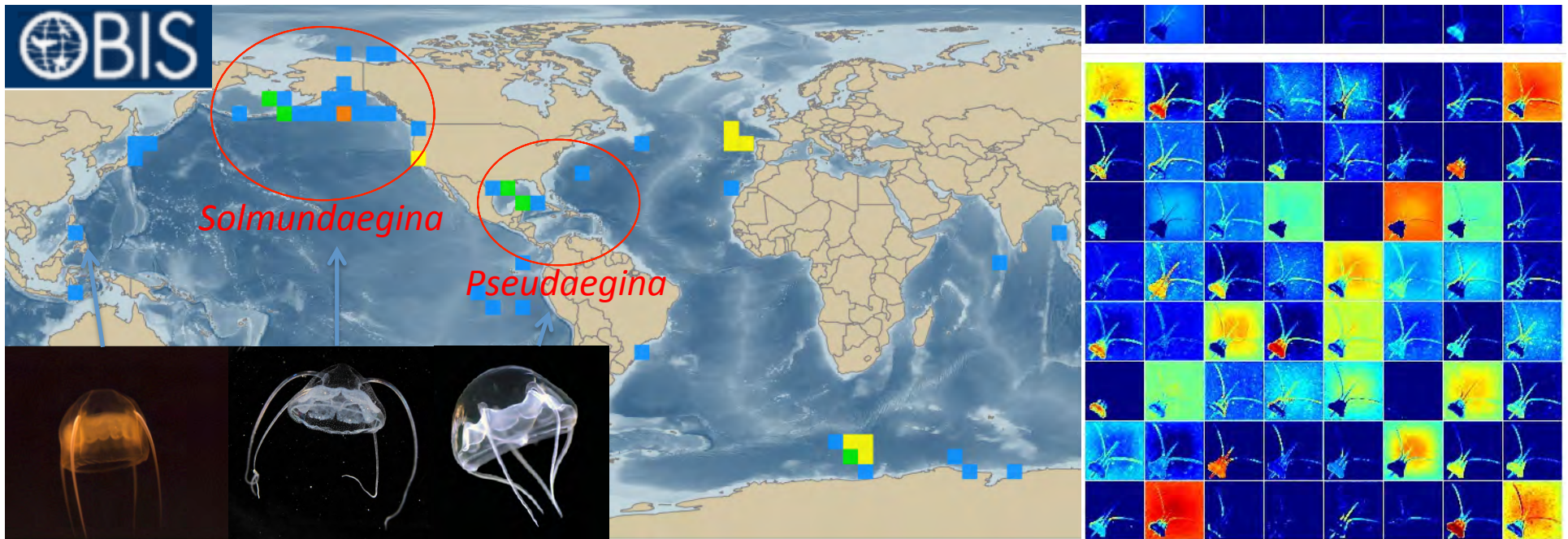


# Recommendations

- Archive original data tables on biogeographic database servers with digital object identifiers (DOIs) to ensure data traceability because we were unable to verify most of the data in OBIS
- Always keep a voucher specimen of whatever you sequence.







## What are their REAL distributions?

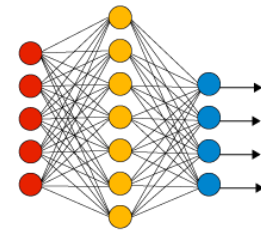
How to check and fix biogeographic database species IDs in light of Access and Benefit Sharing (ABS) policy/legal uncertainty?

Possible to use photos from SCUBA divers and ROV videos to assess?

Too many images for a human to work through...

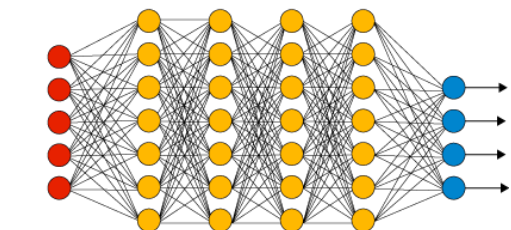
Can a Deep-Learning Machine Vision AI approach be used?

Simple Neural Network



● Input Layer

Deep Learning Neural Network



● Hidden Layer ● Output Layer

# Deep Learning AI

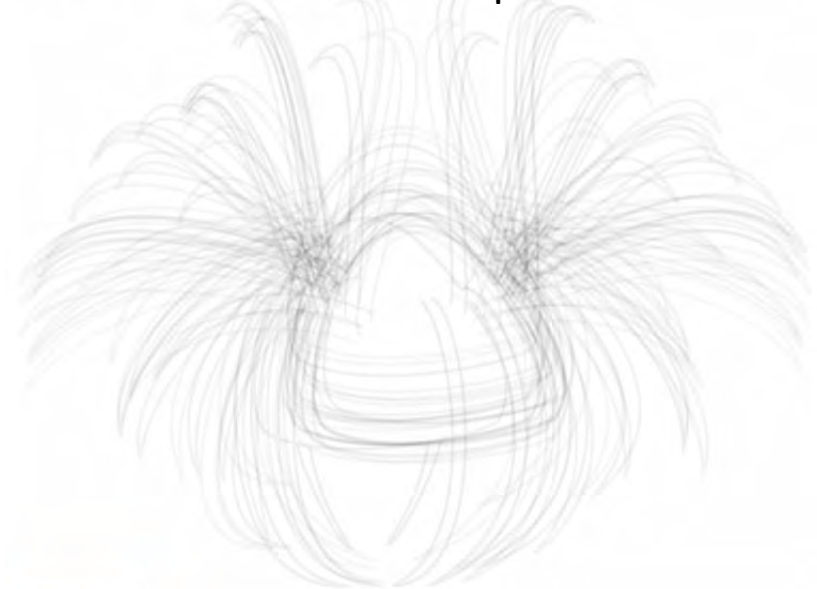
NVIDIA Deep Learning GPU Training System (DIGITS) with CAFFE framework

```
终端
2017-07-14 09:43:13 [INFO ] Loaded 47 jobs.
2017-07-14 09:43:13 [WARNING] Failed to load 3 jobs.
2017-07-14 09:43:13 [DEBUG] 20170628-093457-22a5 - IOError: [Errno 2] No such file or directory: '/usr/local/DL-Box/digits-4.0/digits/jobs/20170628-093457-22a5/status.pickle'
2017-07-14 09:43:13 [DEBUG] 20170704-105644-afbe - IOError: [Errno 2] No such file or directory: '/usr/local/DL-Box/digits-4.0/digits/jobs/20170704-105644-afbe/status.pickle'
2017-07-14 09:43:13 [DEBUG] 20170704-105938-97f4 - IOError: [Errno 2] No such file or directory: '/usr/local/DL-Box/digits-4.0/digits/jobs/20170704-105938-97f4/status.pickle'

D I G I T S 4.0.0

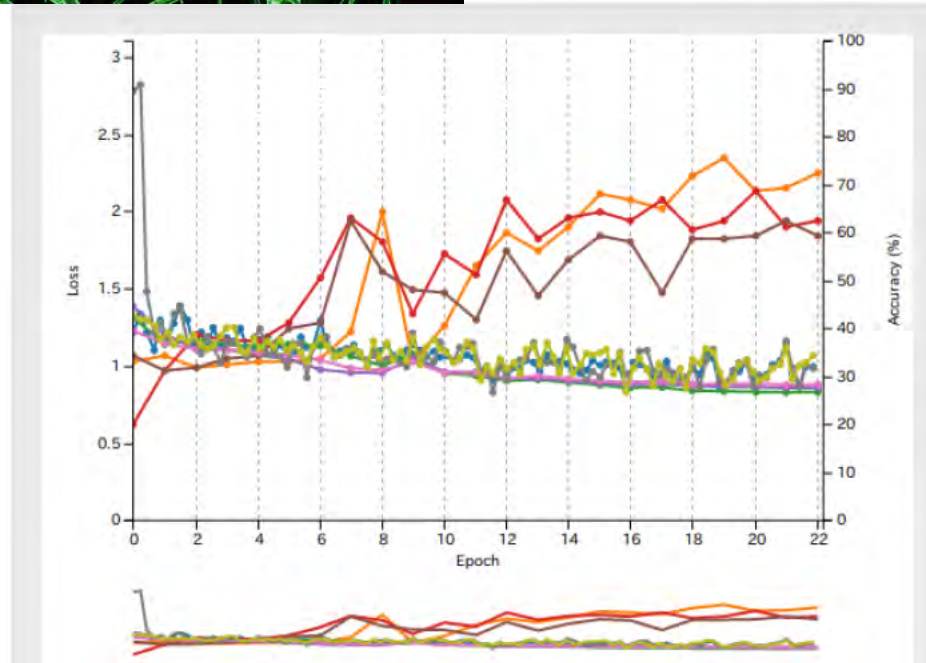
* Running on http://0.0.0.0:34448/
127.0.0.1 - - [2017-07-14 09:43:19] "GET / HTTP/1.1" 200 29666 0.100632
127.0.0.1 - - [2017-07-14 09:43:19] "GET /static/css/bootstrap.min.css HTTP/1.1" 304 190 0.007769
127.0.0.1 - - [2017-07-14 09:43:19] "GET /static/js/bootbox.min.js HTTP/1.1" 304 188 0.003822
127.0.0.1 - - [2017-07-14 09:43:19] "GET /static/css/bootstrap-theme.min.css HTTP/1.1" 304 189 0.003185
```

Image Mean:  
2,4,6 tentacles at a variety of angles  
as would occur in plankton net samples



[Explore the db](#)

Choice of network?  
GoogleNet [22 layers] gave better results than AlexNet [8 layers] (but still not great..)  
70-80% accuracy



## GPU Usage

TITAN X (Pascal) (#0)

**Memory**  
3.85 GB / 11.9 GB (32.4%)  
**GPU Utilization**  
96%  
**Temperature**  
50 °C

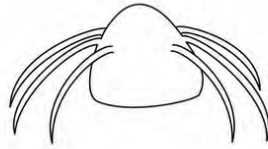
### Related jobs

Image Classification Dataset

NarcomedusaTest Done



# honbantestgoogle1 Image Classification Model



## Predictions

DeepLearning Aegina rosea	54.1%
DeepLearning Aegina citrea	45.28%
DeepLearning Solmundella bitentaculata	0.61%

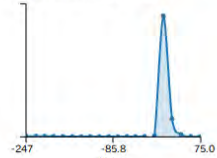
## Description

data

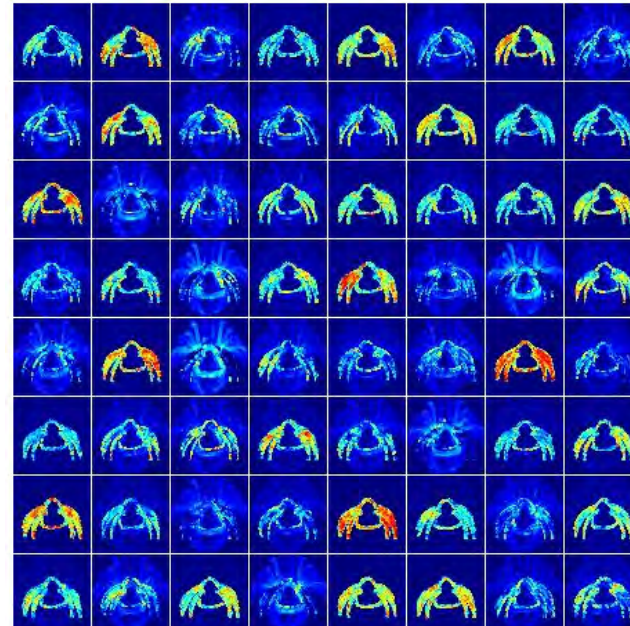
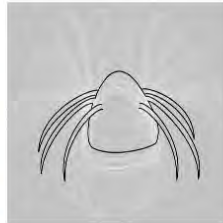
Activation

## Statistics

Data shape: [ 1 224 224]  
Mean: -1.67496  
Std deviation: 33.4134

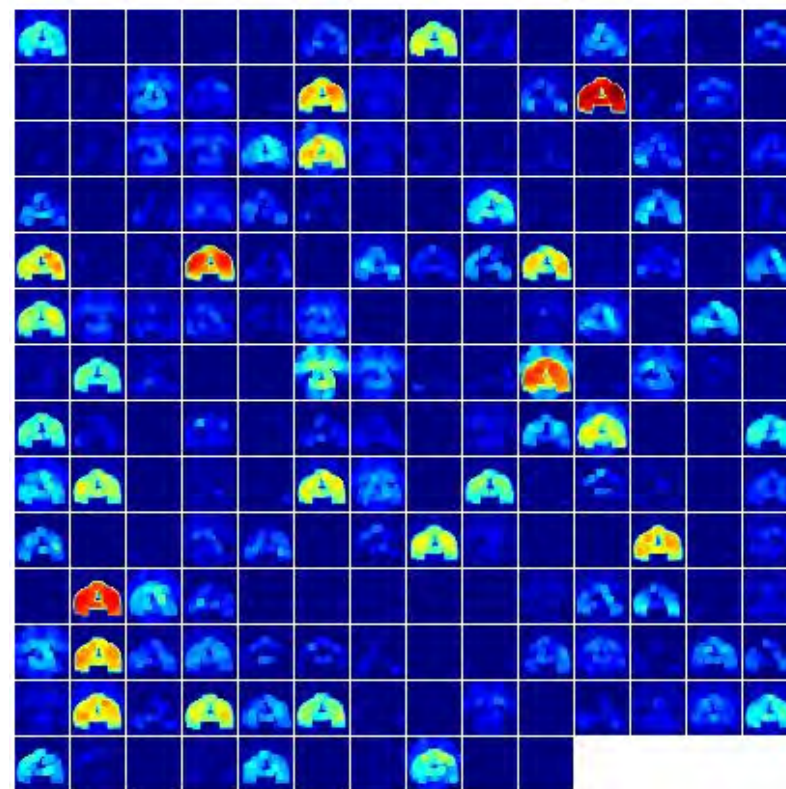
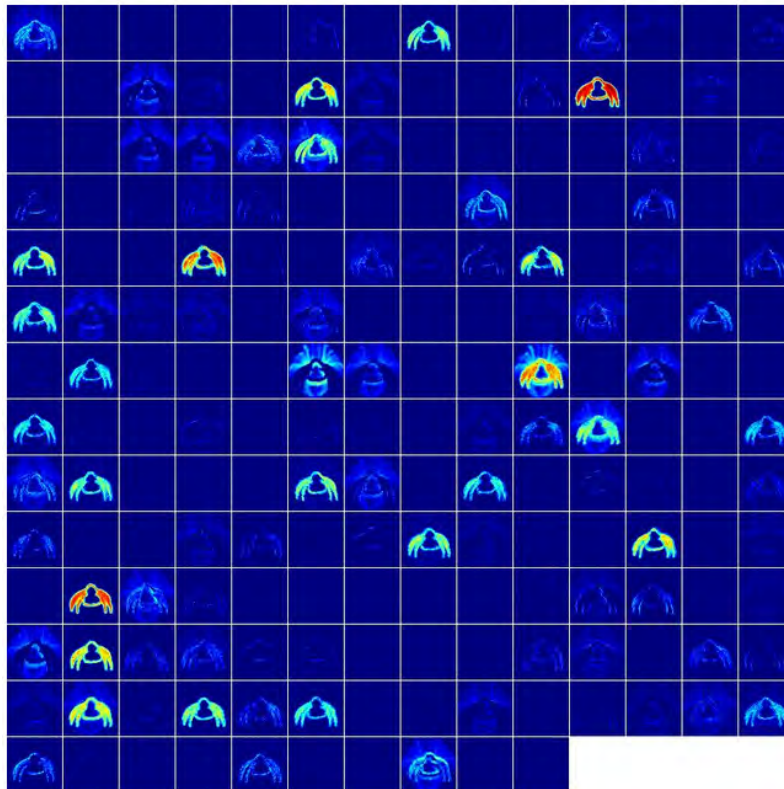


## Visualization



tentacle integrity only up to second convolution

Not-so-Deep Learning better?



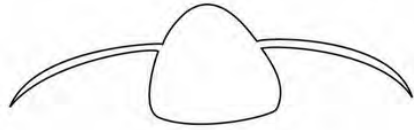


# honbantestgoogle1 Image Classification Model

59% correct

Predictions

DeepLearning Solmundella bitentaculata	59.34%
DeepLearning Aegina citrea	24.16%
DeepLearning Aegina rosea	16.49%



Description

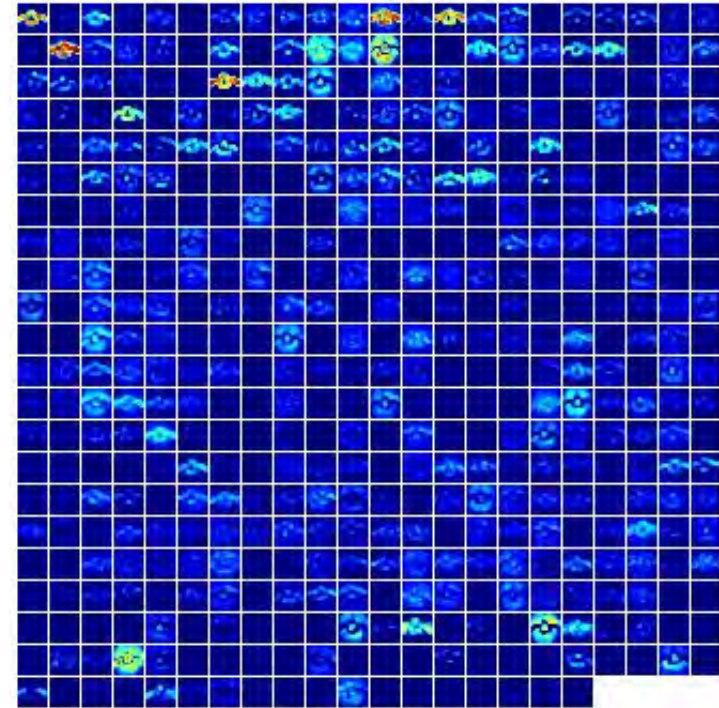
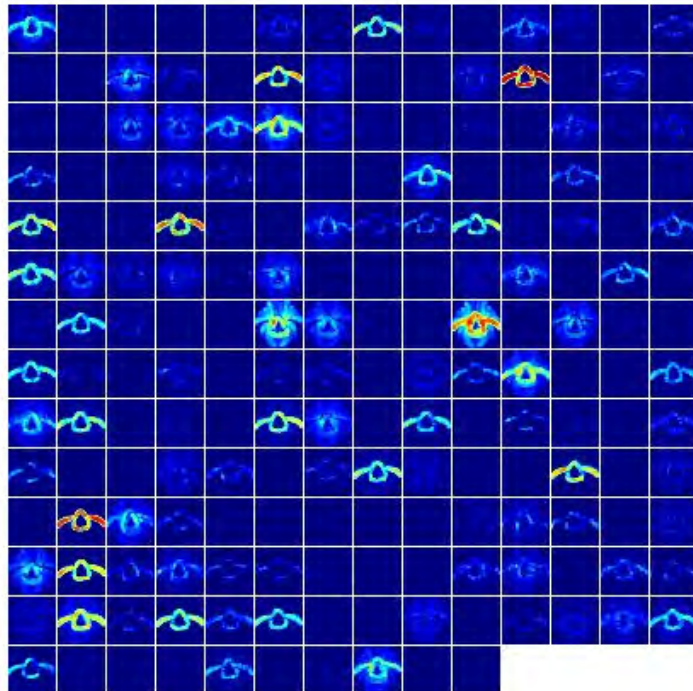
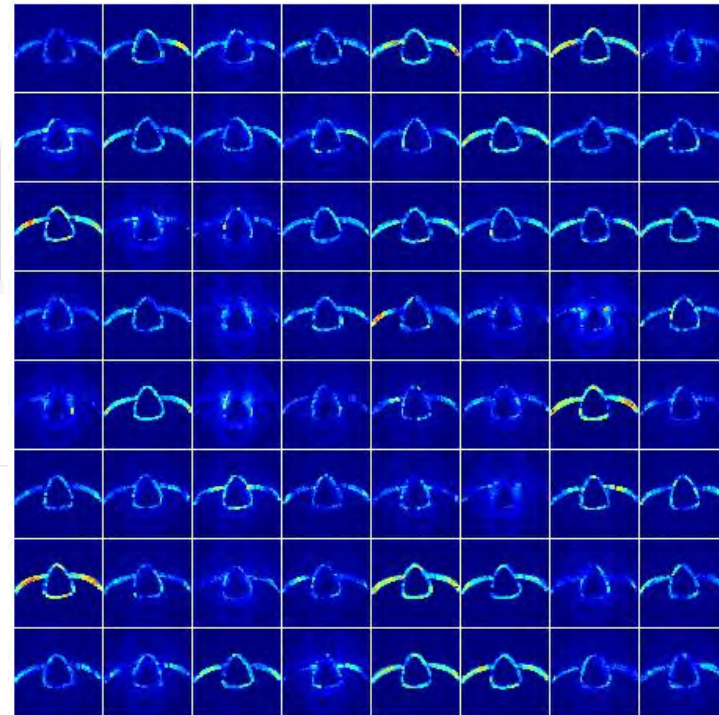
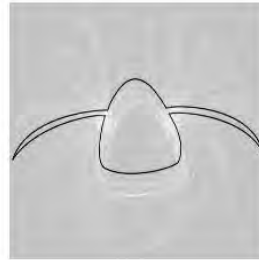
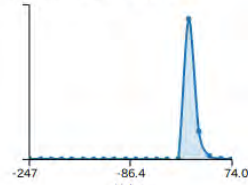
Statistics

Visualization

data

Activation

Data shape: [ 1 224 224]  
Mean: 1.75593  
Std deviation: 24.6611





Made an in situ image training set for 2-6 tentacle Narcomedusae using framegrabs from JAMSTEC ROV & HOV videos convert videos to still frame series, extract jellies manually with proprietary software, flip and rotate images to increase number of images in training set

The screenshot shows a Mozilla Firefox browser window displaying a dataset of Narcomedusae images. The browser address bar shows the URL: `0.0.0.0:34448/datasets/images/classification/explore?db=train_db&job_id=201`. The page title is "DIGITS Image Classification Dataset". The main content area displays a grid of images of Narcomedusae, with labels "Ac", "Bk", "Sb", and "Ag" below them. To the right of the browser window, a terminal window shows logs of HTTP requests and responses, including paths like `/datasets/images/classification/explore` and `/jobs/20170801-133534-fe78`. Below the terminal window, there are two images of a white ROV (L'AAR 6500) and a blue ROV (HYPERDOL) on the water surface. The background of the slide features a green, glowing, tentacle-like pattern. The text "Learning BOX" is visible in the bottom right corner of the slide.

500-700 images per species

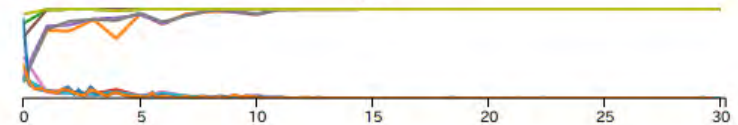
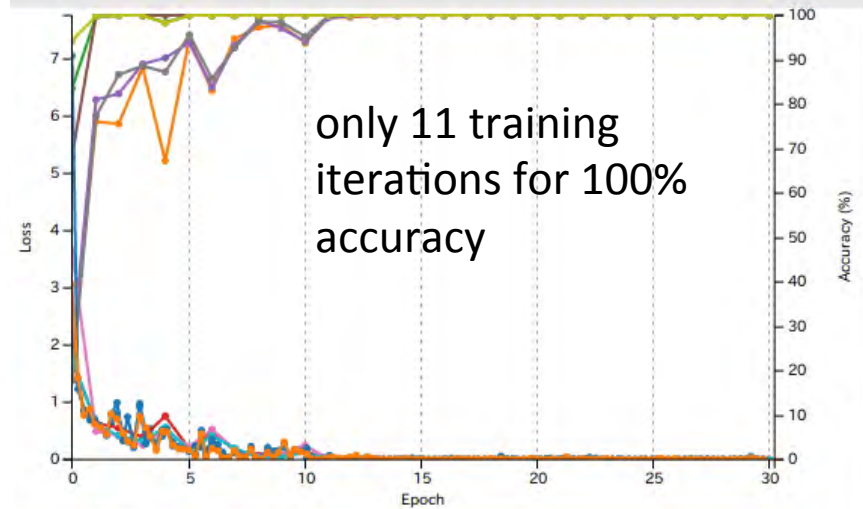
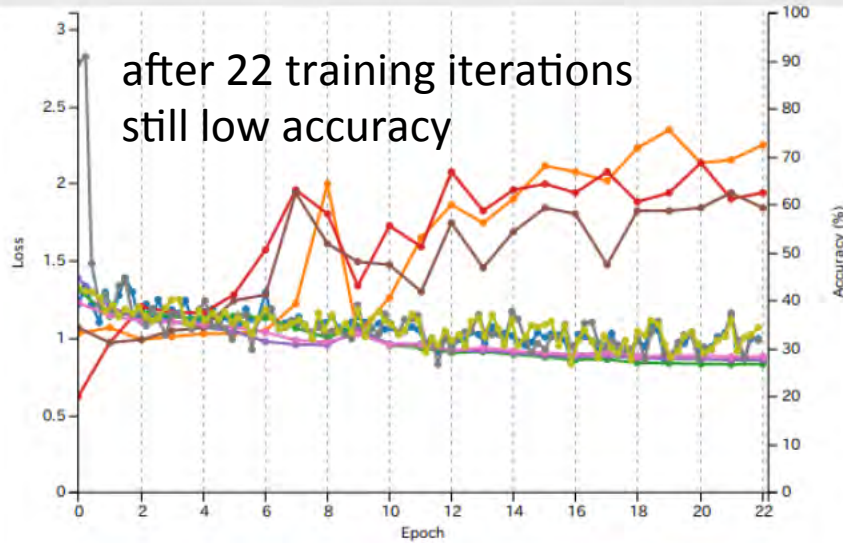
Image Mean:

for humans line drawings  
look more like jellyfish



Explore the db

for Deep Learning this worked better!





# Tried with “real” *Aegina citrea* framegrabs. Classified image from internet with **99.9% accuracy!!**

When alive is yellow, always has tentacles forward and not touching each other, tentacles same length vs bell height.

The screenshot shows a desktop environment with a Firefox Web Browser window open. The browser displays a classification interface for a jellyfish image. The interface includes a file explorer on the left, a terminal window at the top left, and a web browser window on the right. The web browser window shows the classification results for the image, including a prediction table, a description, statistics, and visualization.

**File Explorer:** Shows a directory structure with files and folders. A blue arrow points to a file named 'jellyfish' in the 'examples' folder.

**Terminal Window:** Shows a command prompt with the following output:

```
2017-08-01 14:19:41
2017-08-01 14:19:41
2017-08-01 14:19:41
127.0.0.1 - - [2017
_one?job_id=2017080
2017-08-01 14:21:44
2017-08-01 14:21:44
[2017-08-01 14:21:44] time task subprocess
bin/python? /usr/local/DL-Box/digits-4.0/tools/inference.py
```

**Web Browser:** The browser window shows the classification interface for the image. The URL is `0.0.0.0:34448/models/images/classification/classify_one?job_id=20170801-112`. The interface includes a prediction table, a description, statistics, and visualization.

**Predictions:**

Class	Accuracy
Ac	99.88%
Sn	0.12%
Ag	0.0%
Bk	0.0%
Ar	0.0%

**Description:**

data  
Activation

**Statistics:**

Data shape: [ 3 224 224]  
Mean: -38.8417  
Std deviation: 37.7548

**Visualization:**

conv1/7x7\_s2  
Weights (Convolution layer)  
9,472 learned parameters

Data shape: [ 64 3 7 7]  
Mean: 0.000119078  
Std deviation: 0.0880761

The visualization shows a heatmap of the convolution layer weights, with a blue arrow pointing to the Google logo in the background image.

# *Aeginura grimaldii* also 99.9% accuracy!!

When alive is brown-red and tentacles always sideways or down.

The screenshot displays a Jupyter Notebook environment. On the left, a terminal window shows a warning: `[WARNING] Infer Model unrecognized output: /usr/local/DL-Box/digits-4.0/digits/util/image.py:108: VisibleDeprecationWarning: using a non-integer number instead of an integer will result in an error in the future`. Below the terminal is a file browser showing a grid of image thumbnails. A blue arrow points from the terminal to a specific image. On the right, the 'Classify One Image' web application is open, showing a classification result for a brown-red organism. The 'Predictions' table is as follows:

Class	Accuracy
Ag	99.9%
Ar	0.08%
Ac	0.01%
Sn	0.0%
Bk	0.0%

The application also displays a histogram for the 'data' activation, a visualization of the 'conv1/7x7\_s2' weights, and a Google search bar with the text 'Aegina citrea'.



When the 2-tentacled *Solmundella bitentaculata* was not in the training set it was **misidentified** as *Solmundagina nematophora* (same family) **at 94% certainty** level.

Need to ensure training set contains all species!!!

The screenshot shows a web browser window with the URL `0.0.0.0:34448/models/images/classification/classify_one?job_id=20170801-112`. The page displays the results of a classification task for a nematode image. The predicted species and their confidence scores are:

Species	Confidence
Sn	93.68%
Ac	3.55%
Ar	2.27%
Bk	0.31%
Ag	0.2%

The interface also shows a 'Notes' section with 'None' and a 'Visualization' section for the 'conv1/7x7\_s2' layer, which displays a heatmap of the input image with a red 'X' overlaid, indicating a misclassification.

*Solmundaegina nematophora* was correctly identified to **86% certainty** (9% *A. citrea*).

Images do seem useful for automatic generation of species occurrence data from SCUBA photos or ROV images **as a result of this taxonomic work**

Training set image resolution 256x256 pixels (cheap computer...). Better with 512x512 pixels?

The screenshot shows a desktop environment with a terminal window on the left displaying logs for model inference tasks. The terminal output includes timestamps and status messages such as "Infer Model task complete" and "Task subprocess".

In the center, a file explorer window displays a grid of image thumbnails, each with a corresponding filename and file size. The files are organized into folders, and the thumbnails show various images, including what appears to be a jellyfish.

On the right, a web browser window displays the "Classify One Image" interface. The main image shows a jellyfish. The "Predictions" section lists the following results:

Species	Confidence
Sn	86.29%
Ac	6.79%
Ar	2.52%
Bk	1.69%
Ag	0.69%

The "Notes" section is currently empty. Below the main image, there are three sections: "Description", "Statistics", and "Visualization".

The "Statistics" section shows the following data for the "data" layer:

- Data shape: [3 224 224]
- Mean: -58.9718
- Std deviation: 34.7897

The "Visualization" section shows a heatmap of the image, with a blue arrow pointing to a specific region. Below this, there is a search bar with the text "Aegina citrea" and the Google logo.

The "conv1/7x7\_s2" section shows the weights for a convolution layer:

- Data shape: [64 3 7 7]
- Mean: 0.000119078
- Std deviation: 0.0880761
- 9,472 learned parameters

A histogram of the weights is shown below the statistics.

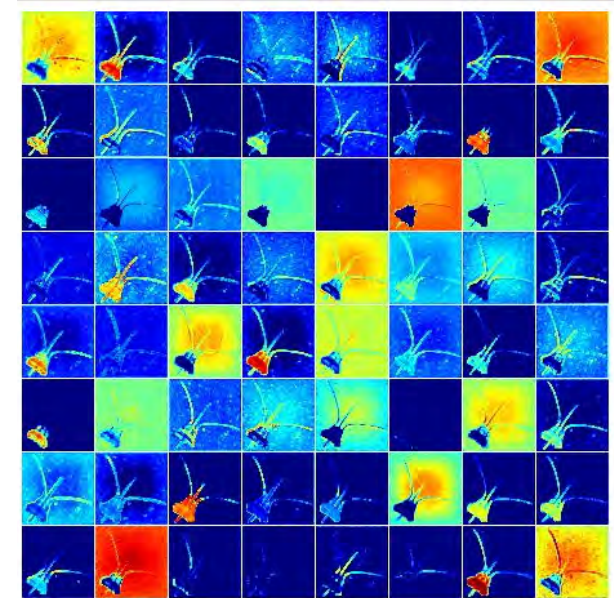
At the bottom right, a text overlay reads: "Quality Control of training set of utmost importance!"



# Summary

If a “set” of in situ imagery and pristine specimens is collected for **taxonomic work** on similar-looking species, it appears that Deep Learning techniques will allow highly accurate classification of that taxonomic group from in situ video imagery.

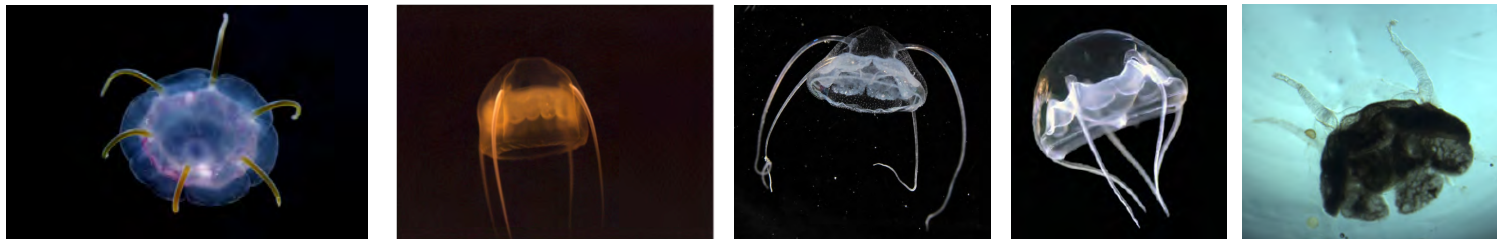
(caveat: only tested with Narcomedusae so far)



accurate biogeographic data  
may be in our sights!

As with GenBank: Rubbish in → Rubbish out

We all know eDNA needs more Groundtruthing  
but so do Deep Learning approaches



# Most up-to-date Taxonomy

Licandro, P., Fischer, A. and Lindsay, D.J. (2017) Cnidaria: Scyphozoa and Non-colonial Hydrozoa. in: Castellani, Edwards, M. (Eds.) Marine Plankton: A Practical Guide to Ecology, Methodology, and Taxonomy. Oxford University Press. 198–231. ISBN 9780199233267.

Licandro, P., Carré, C. and Lindsay, D.J. (2017) Colonial Hydrozoa (Siphonophorae). in: Castellani, Edwards, M. (Eds.) Marine Plankton: A Practical Guide to Ecology, Methodology, and Taxonomy. Oxford University Press. 232–250. ISBN 9780199233267.

Licandro, P. and Lindsay, D.J. (2017) Ctenophora. in: Castellani, Edwards, M. (Eds.) Marine Plankton: A Practical Guide to Ecology, Methodology, and Taxonomy. Oxford University Press. 251–263. ISBN 9780199233267.

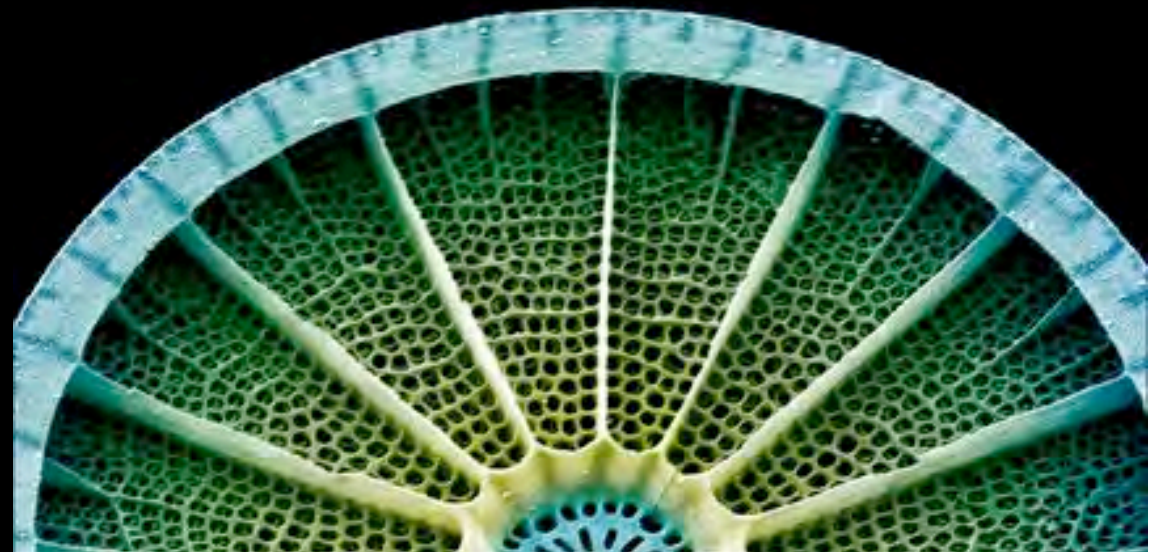


# MARINE PLANKTON

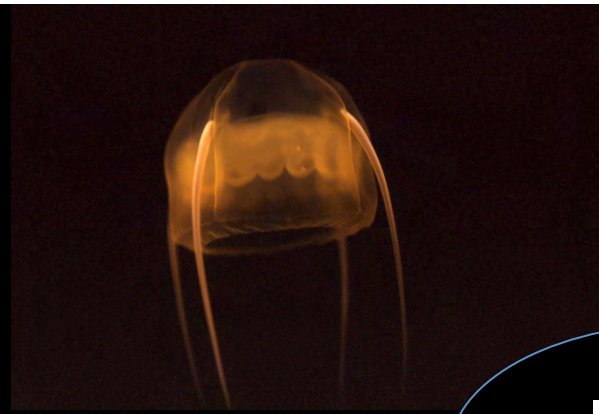
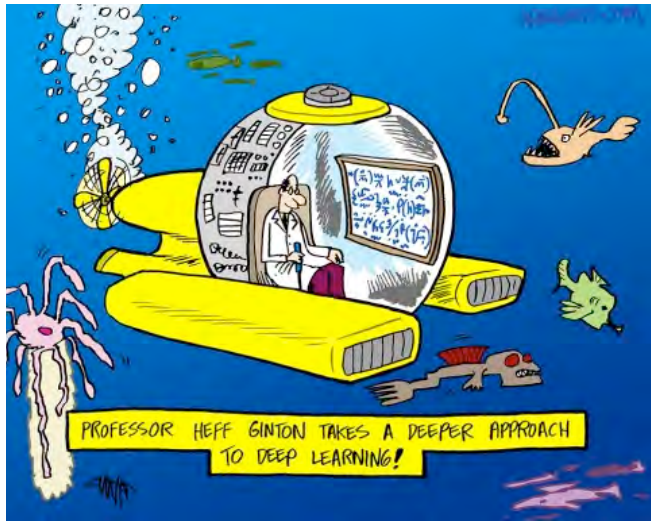
A Practical Guide to Ecology, Methodology, and Taxonomy

*Edited by*

Claudia Castellani & Martin Edwards







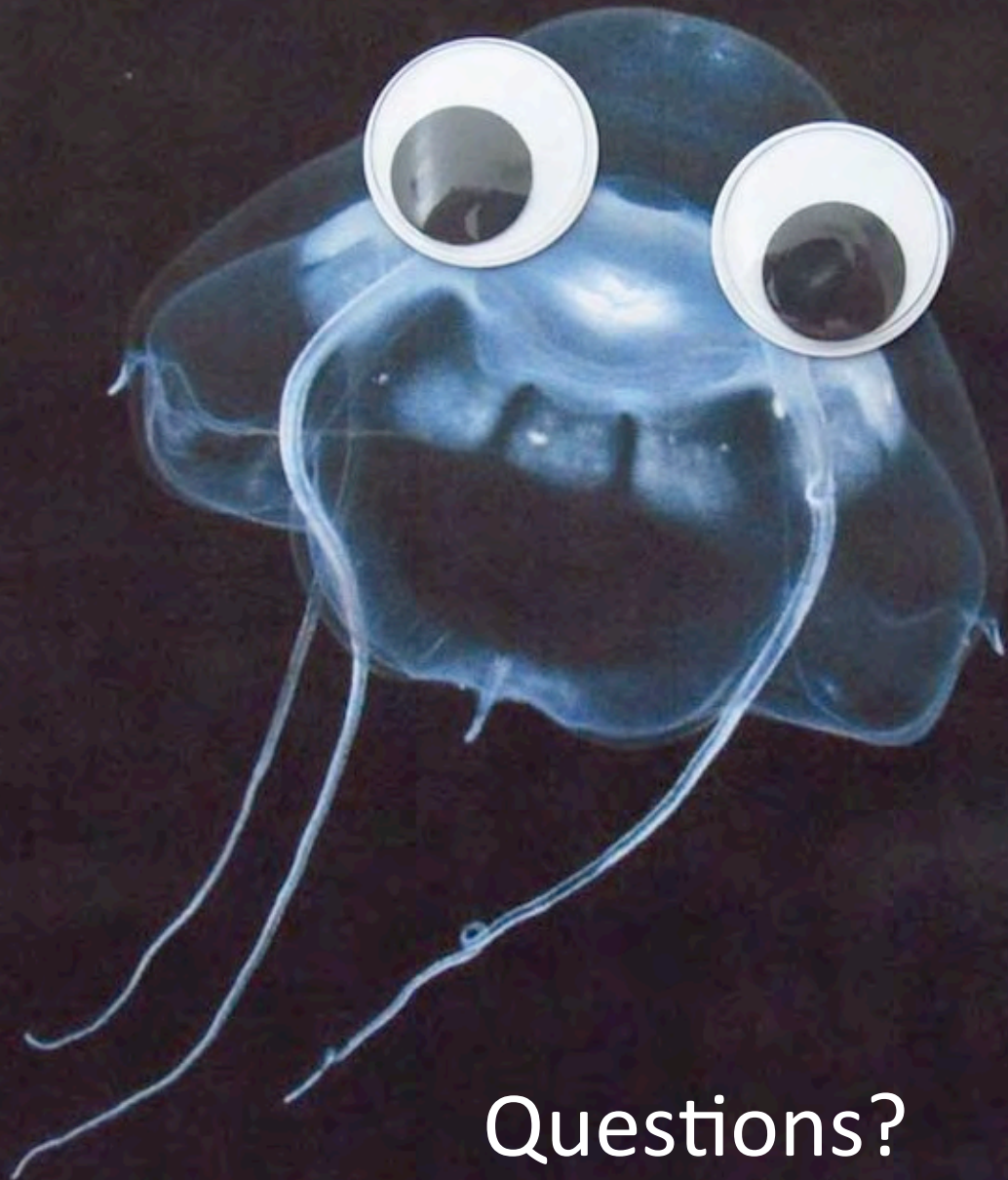
I am not a jellyfish!

Viva!  
good taxonomy  
& Quality Control

Thanks to

Hiroyuki Yamamoto (JAMSTEC)  
Shuhei Nishida (Tokyo U.)  
Susumu Ohtsuka (Hiroshima U.)  
Peter Schuchert (Muséum d'histoire naturelle, Switzerland)  
Census of Marine Life (CoML [CMarZ, ArcOD]) , INDEEP, DOSI

NSF, NOAA, JSPS, CREST, SIP & the crews and operations teams of many platforms



Questions?