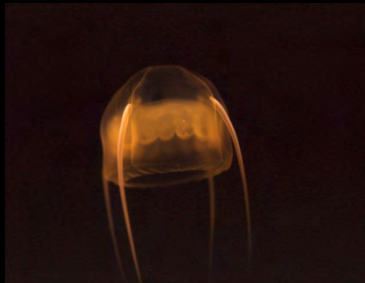# The perils of bad taxonomy for leading edge science: a case study with the genus *Aegina*, and the consequences for Deep Learning

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
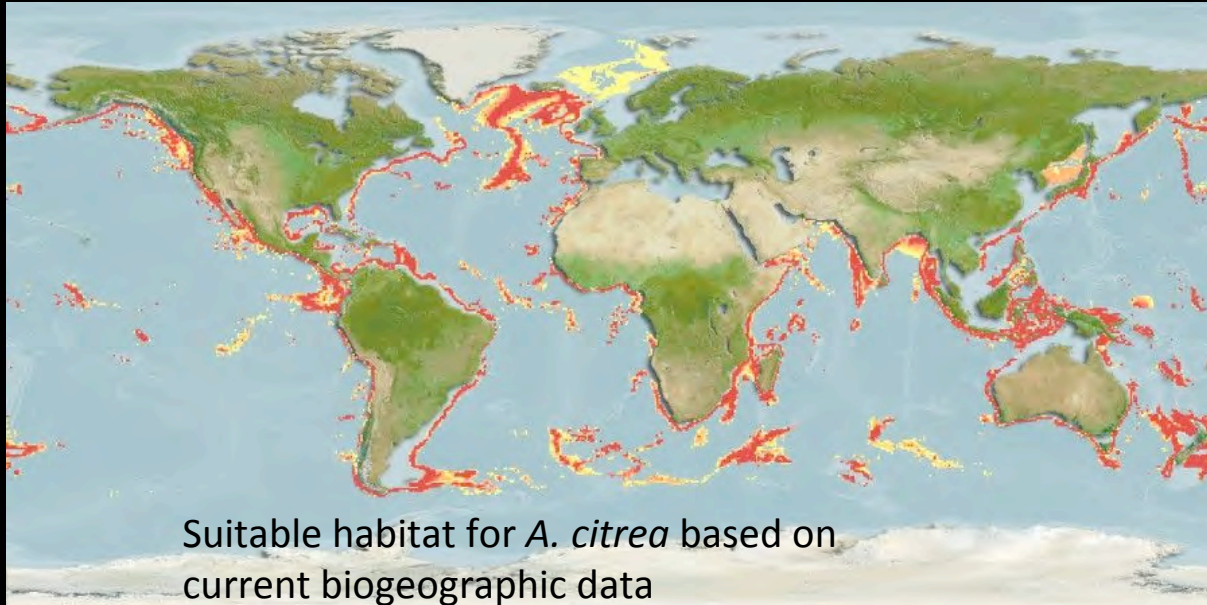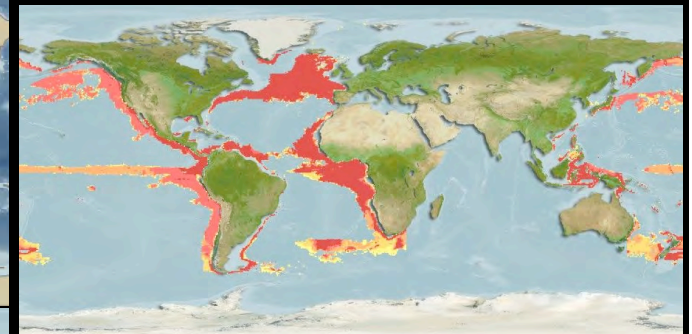all relegated to this one species

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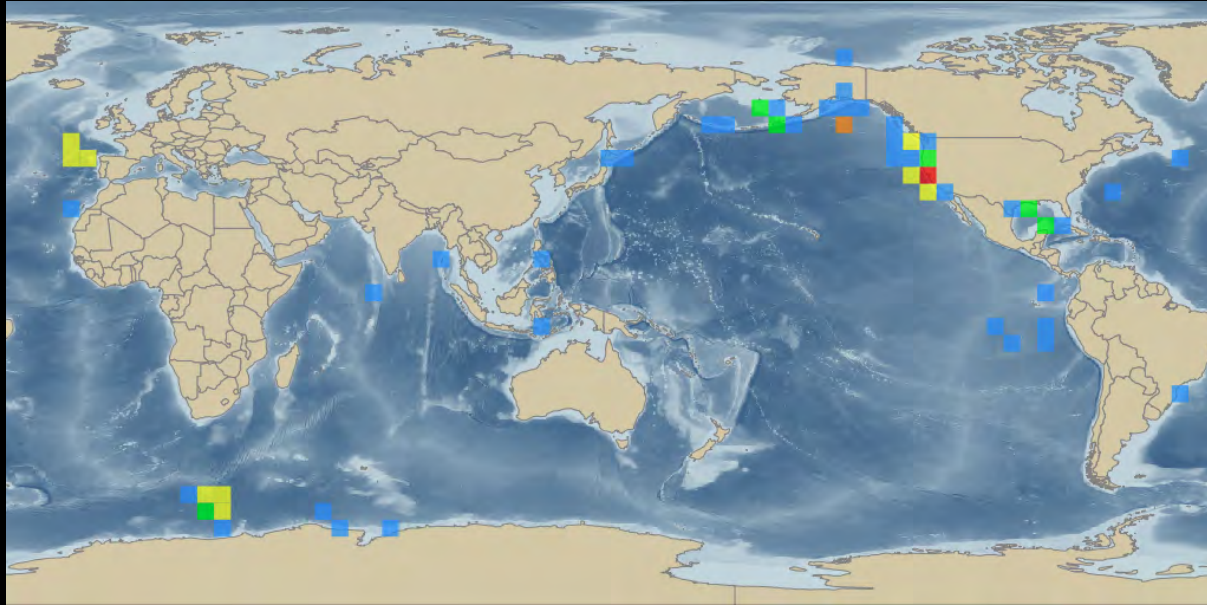
Aegina citrea

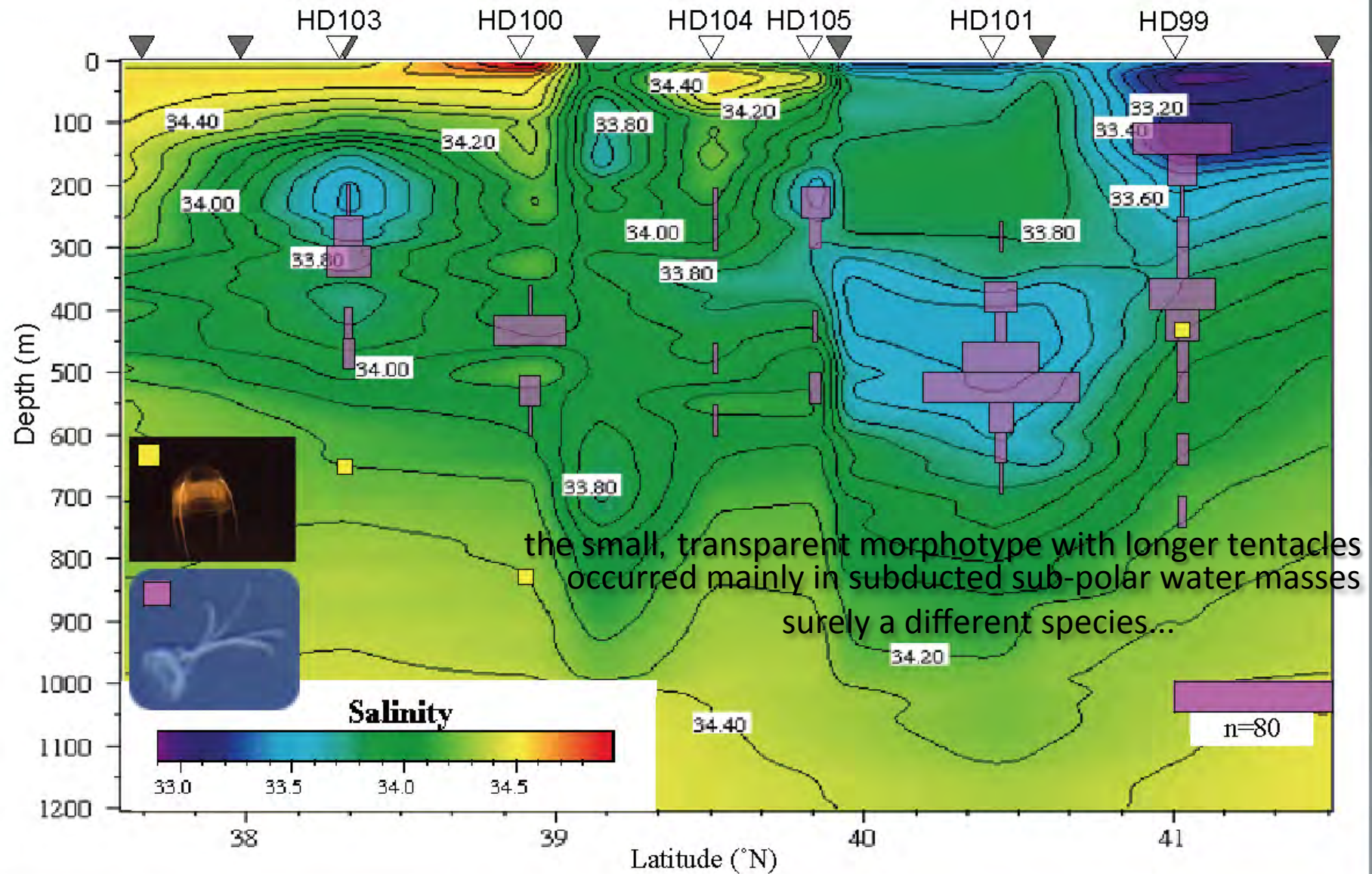What data is in OBIS?

Aegina citrea (Year 2050 range)

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

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[1], BASTIAN BENTLAGE[2], ALBERTO LINDNER[3], DHUGAL LINDSAY[4], STEVEN H.D. HADDOCK[5], GERHARD JARMS[6], JON L. NORENBURG[7], THOMAS JANKOWSKI[8] AND PAULYN CARTWRIGHT[2]
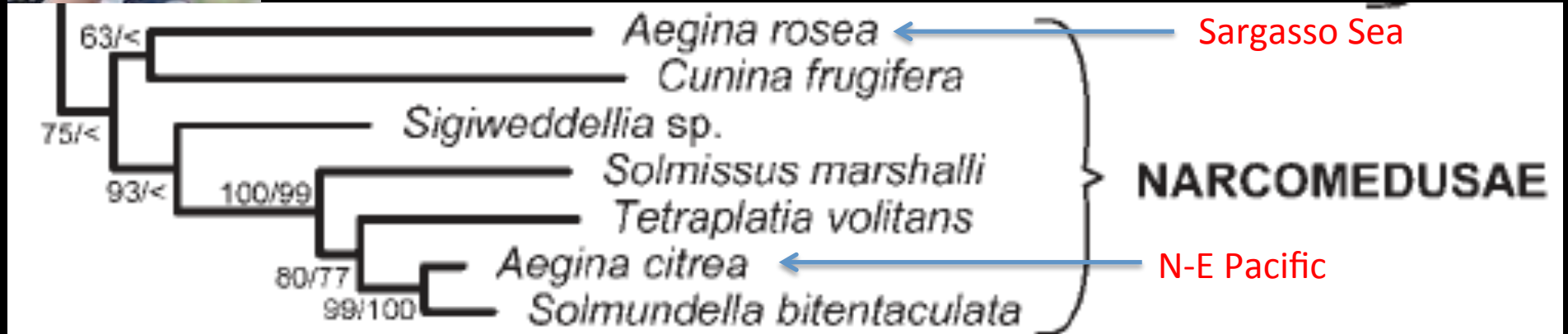
We pointed out in this 2008 paper that *Aegina citrea* and *A. rosea* were genetically different enough to be put in 2 separate families

our "*A. rosea*"

our "*A. citrea*"

Aegina rosea ← Sargasso Sea
Cunina frugifera
Sigiweddellia sp.
Solmissus marshalli
Tetraplatia volitans
Aegina citrea ← N-E Pacific
Solmundella bitentaculata

NARCOMEDUSAE

63/<
75/<
93/<
100/99
80/77
99/100

However to our dismay...

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

GenBank: EU247813.1

Aegina rosea small subunit 18S ribosomal RNA gene

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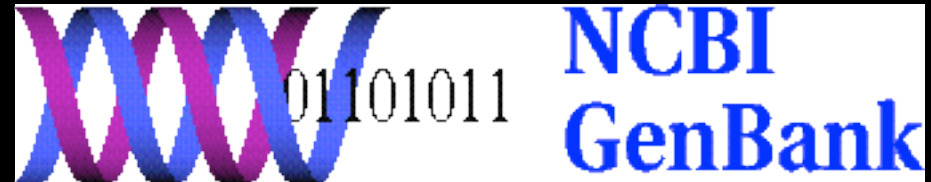
somehow our *Aegina rosea* sequence has been labelled *Aegina citrea*, even though our paper described family level differences and we registered it as *Aegina rosea*!!!

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

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 |

in ITIS *Aegina* contains 3 valid species

3 valid species!

Aegina citrea from type locality (off Japan)

ITIS taxonomy is based on the NODC Taxonomic Code

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

"Aegina rosea" of Collins et al 2008

following Bigelow, 1913

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

GenBank must
follow this

OBIS

WoRMS

OK. We all know GenBank contains mistakes..
What about biogeographic databases?

Biogeography of *Aegina citrea* according to the two biggest biogeographic databases

ITIS

4px

Global Biodiversity
Information Facility
Free and Open Access to Biodiversity Data

our *A. rosea* was from here

ITIS

Aegina rosea
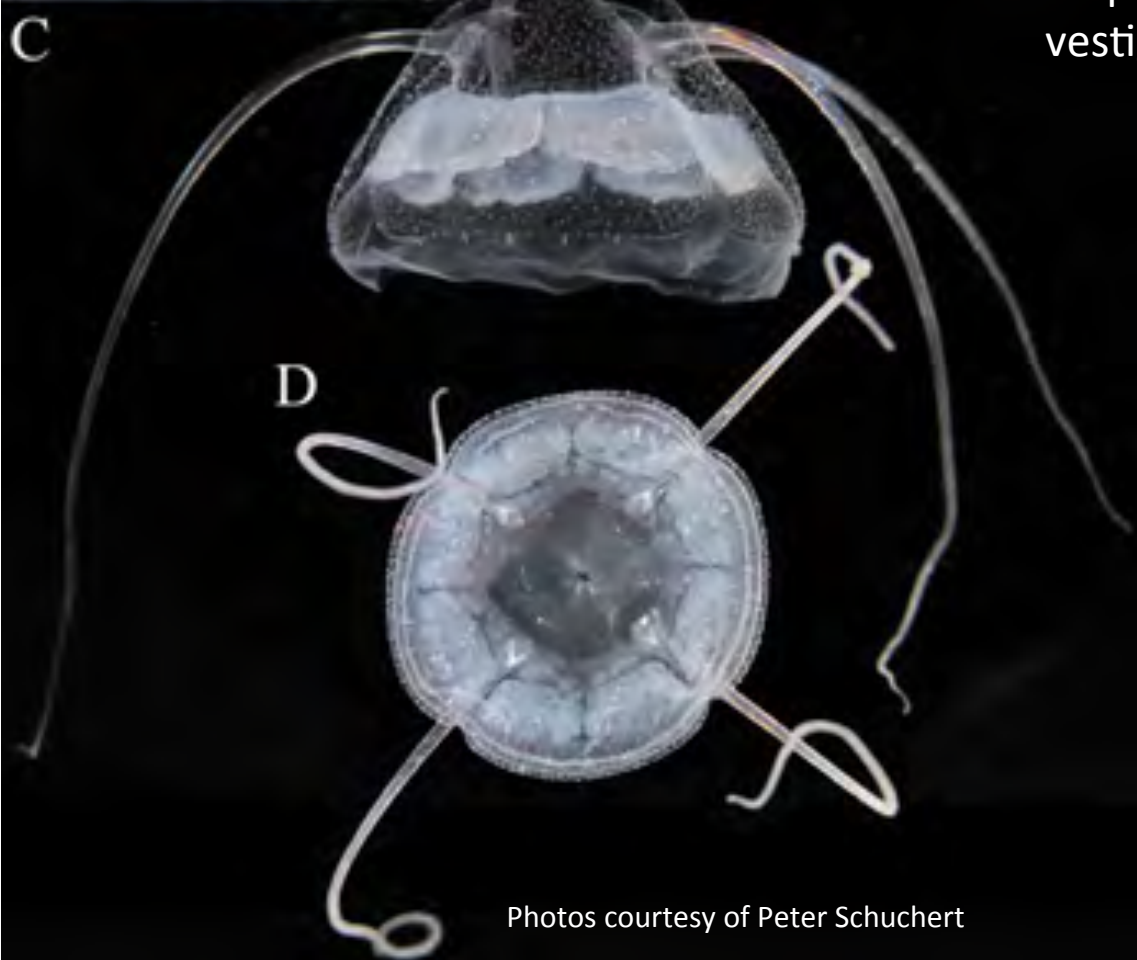
ITIS

Aegina citrea
(incl. synonymized *Aegina rosea*)

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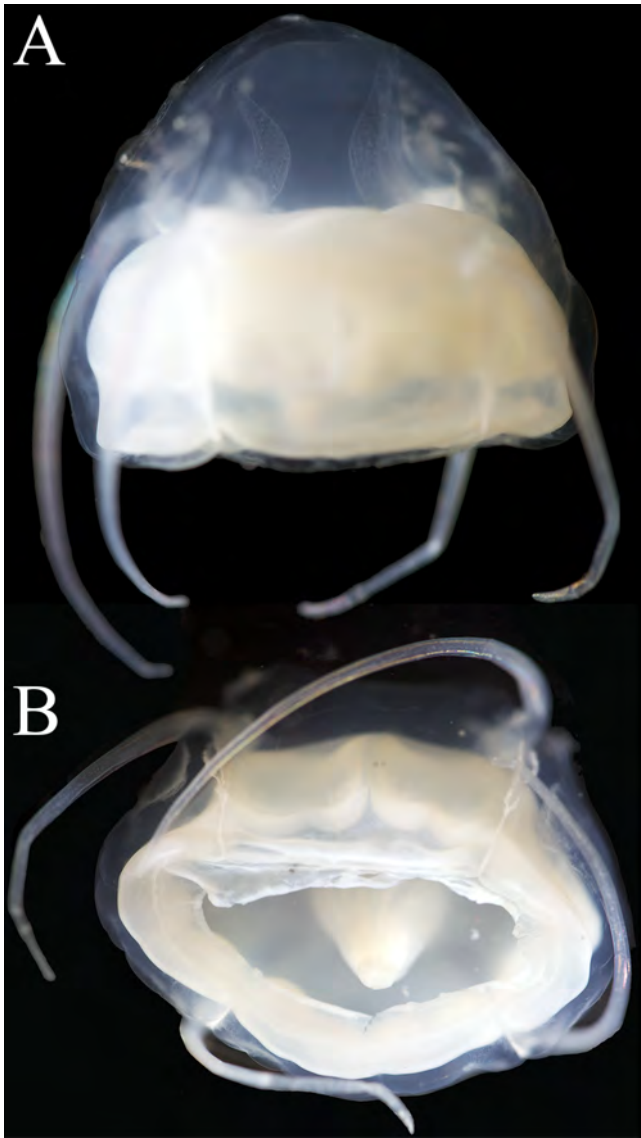*Solmundaegina* Lindsay, 2017
Solmundaegina nematophora Lindsay, 2017

NEW SPECIES, GENUS & FAMILY!!

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

Photos courtesy of Peter Schuchert
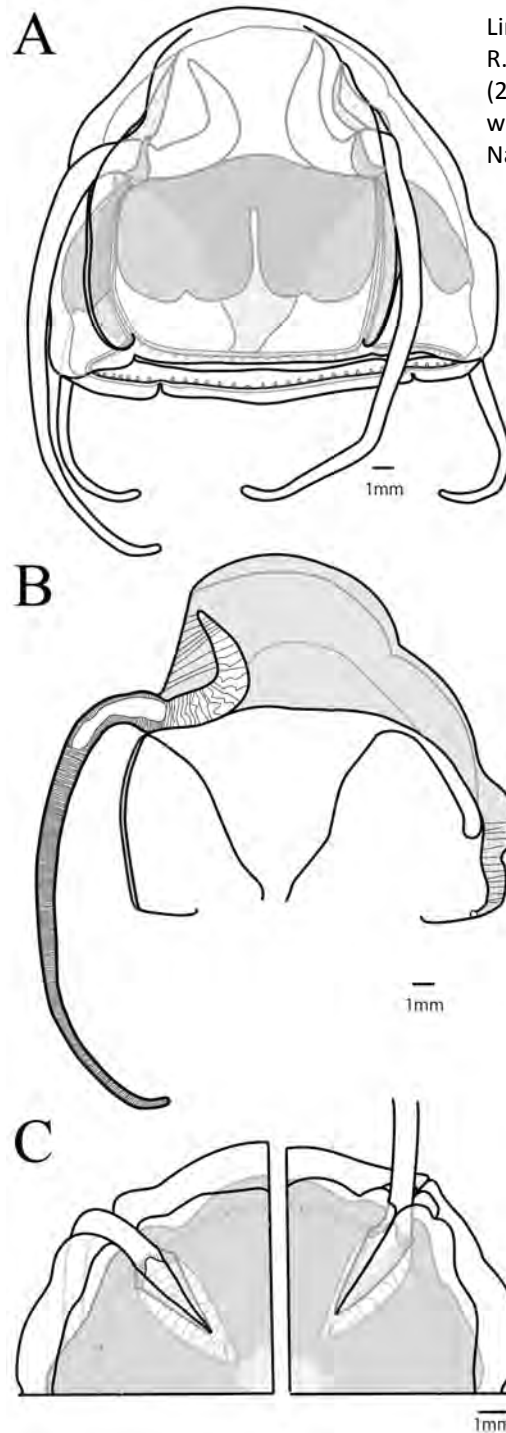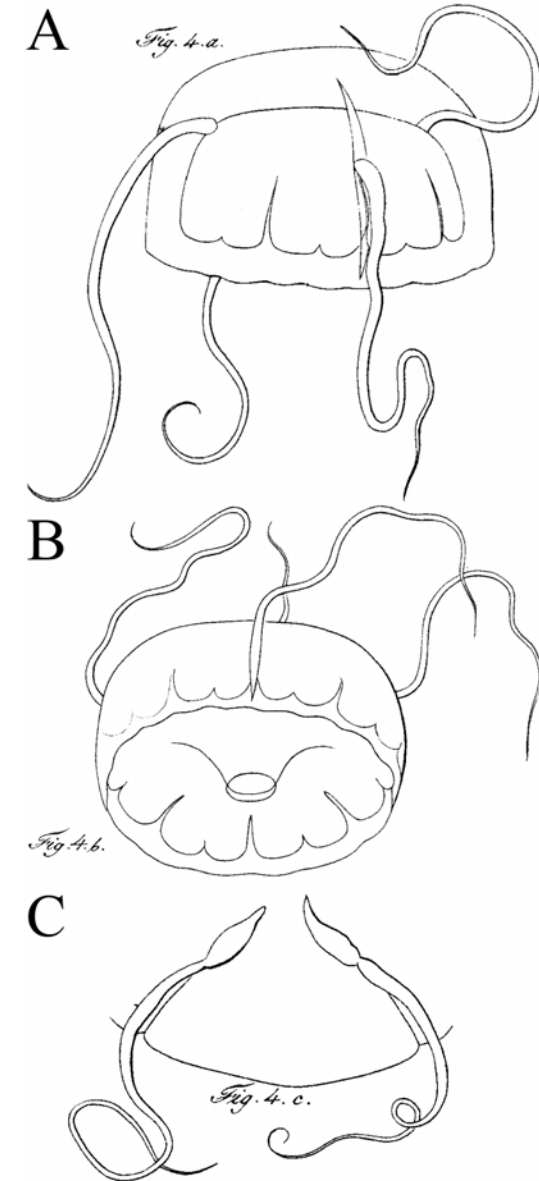
The REAL *Aegina citrea*!
from near the Type Locality off eastern Japan

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

The Collins et al 2008 sequence of "Aegina citrea" is actually the species *Solmundaegina nematophora* (in a different family)! Sorry....

What about the "Aegina rosea" of Collins et al 2008?

100 86 *Aeginona brunnea*
             Aegina citrea [1]
             Aegina citrea [2]
        94 *Aegina citrea* [3]
             Aegina citrea [4]
        86 *Aegina citrea* [5]
             Aegina citrea [6]

Aeginidae

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

100 *Tetraplatia volitans*
        Tetraplatia chuni [1]
100 *Tetraplatia chuni* [2]

Tetraplatidae

65 100 *Solmundaegina nematophora* [1]
            Solmundaegina nematophora [2]
      99 100 *Solmundella bitentaculata* [1]
                  Solmundella bitentaculata [2]
            84 *Aeginopsis laurentii* [1]
                 Aeginopsis laurentii [2]
            100 *Aeginopsis laurentii* [3]
                 Aeginopsis laurentii [4]
                 Aeginopsis laurentii [5]

Solmu

small downward-pointing tentacle roots
no peripheral canal system
vestigial secondary tentacle bulbs
shallow apical grooves
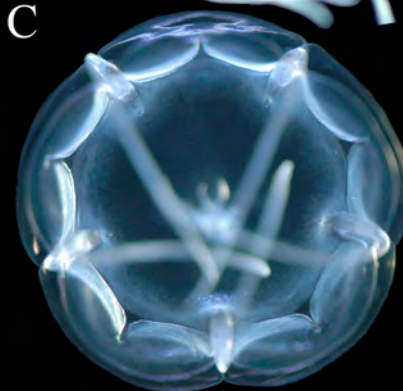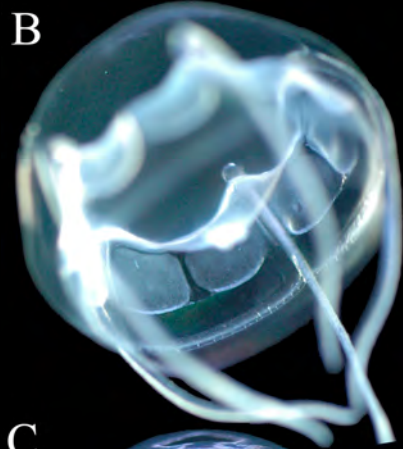
(ML topology based on combined 18S, 16S and COI data)

Pseudaeginidae Lindsay, Bentlage & Collins, 2017
Pseudaegina Lindsay, 2017

Pseudaegina rhodina (Haeckel, 1879)
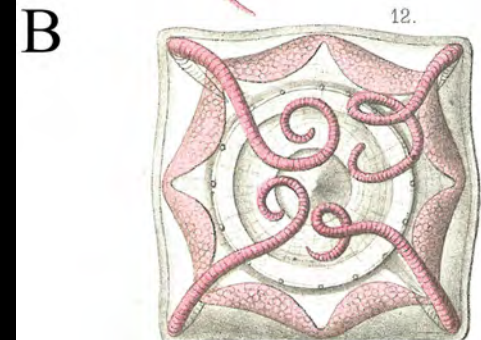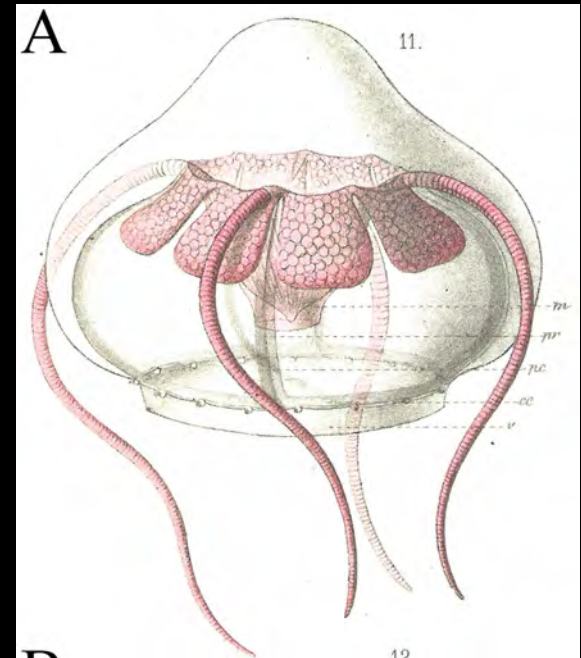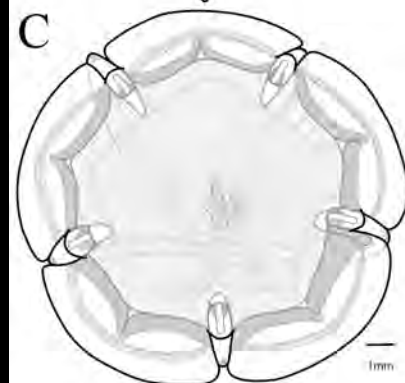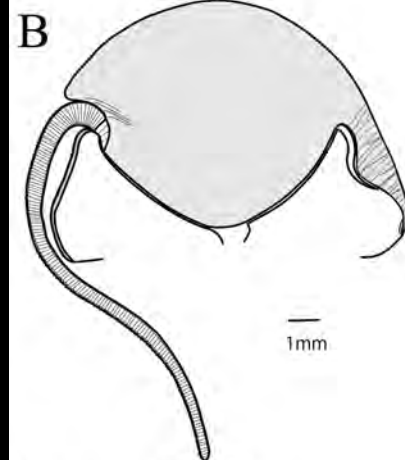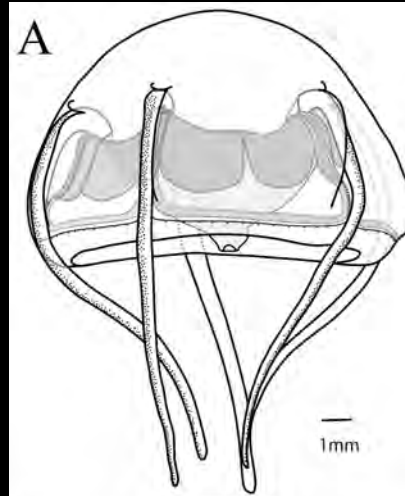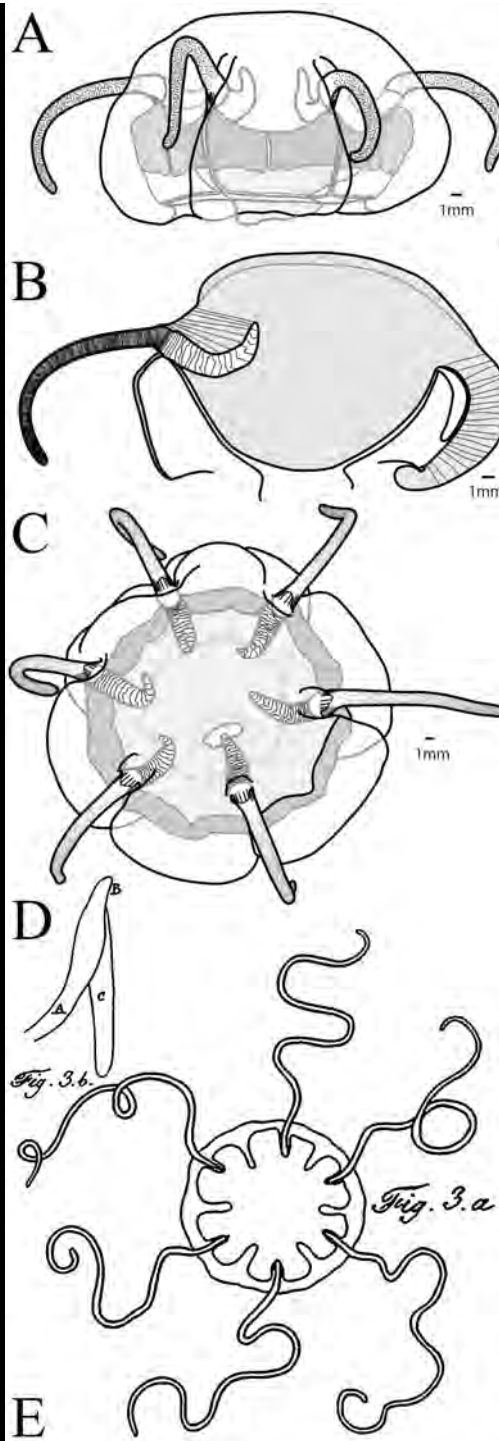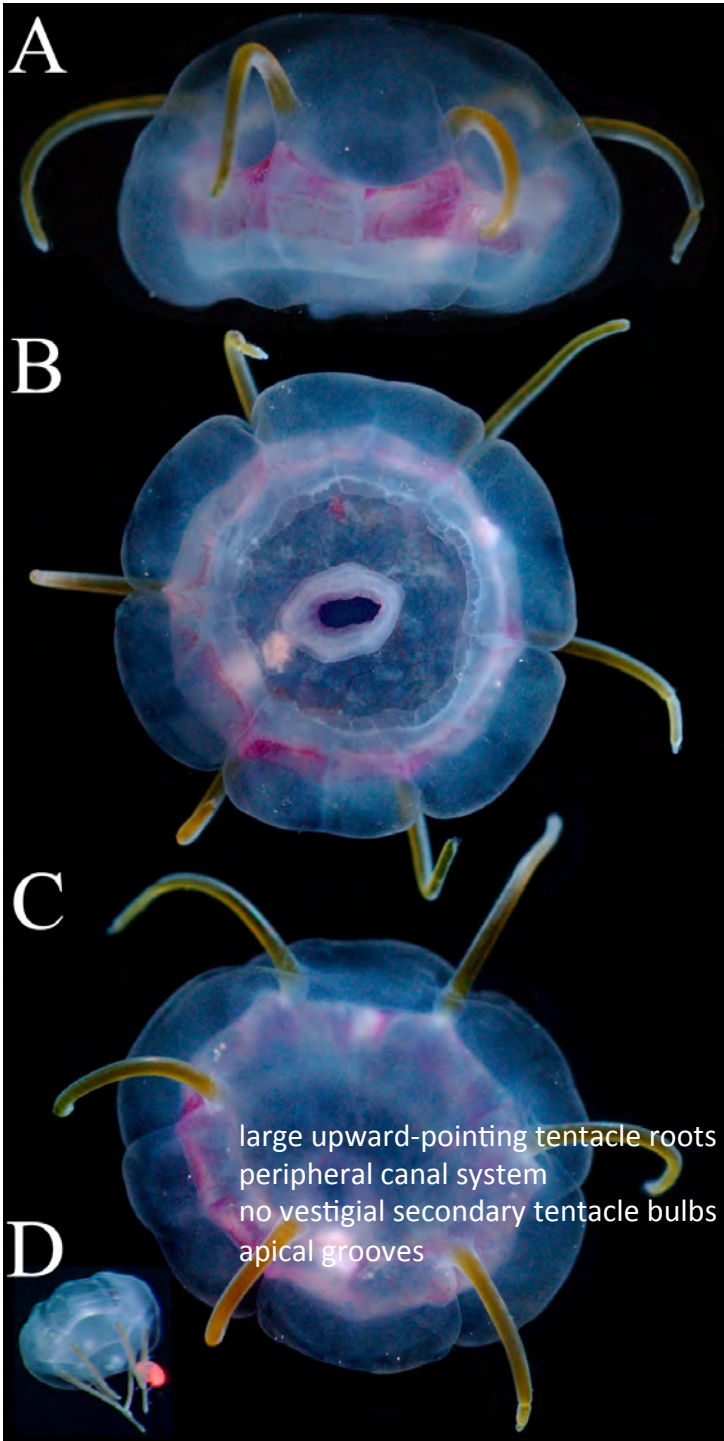
NEW GENUS & FAMILY!!

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
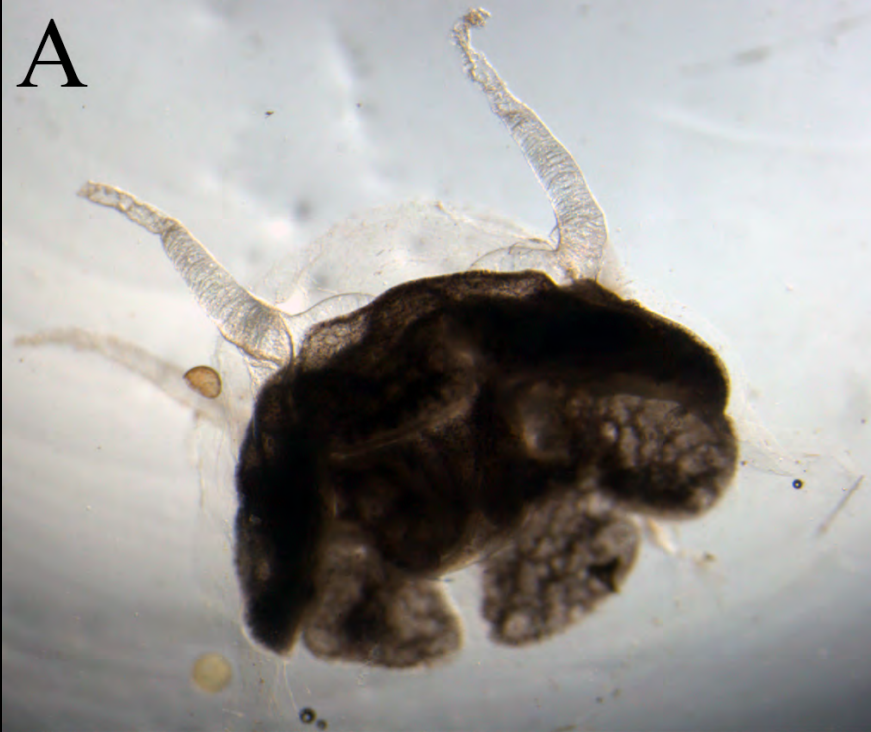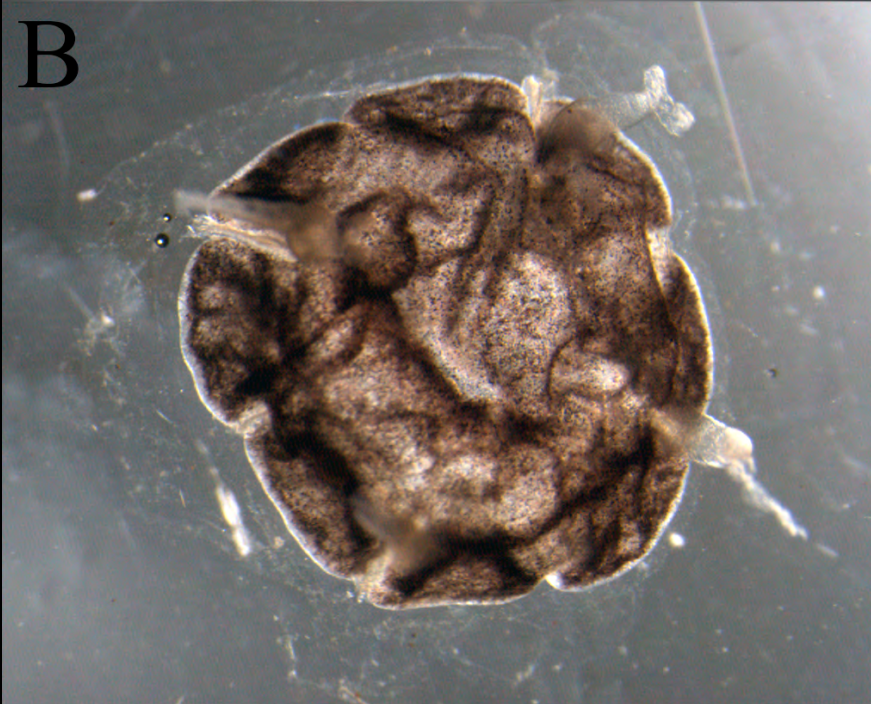orally bulging mesoglea

The REAL *Aegina rosea*!

Not this!

large upward-pointing tentacle roots
peripheral canal system
no vestigial secondary tentacle bulbs
apical grooves
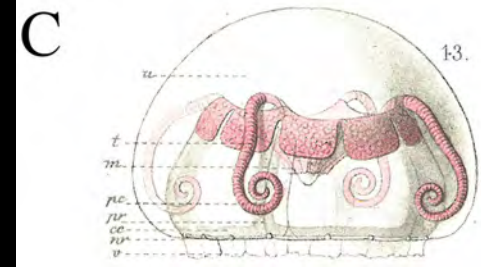
Therefore neither of the Collins et al 2008 sequences of (according to GenBank) "Aegina citrea" are actually that species. Furthermore, the species we labelled *A. rosea* actually belongs to a different family...
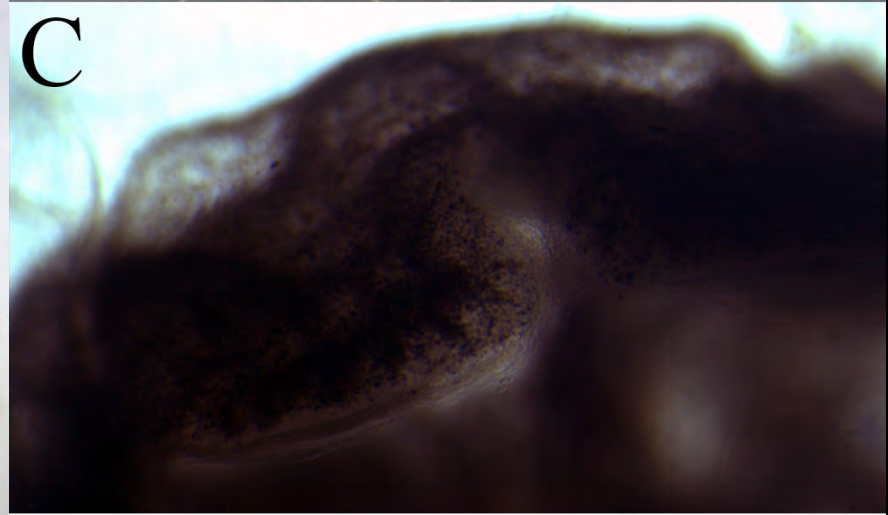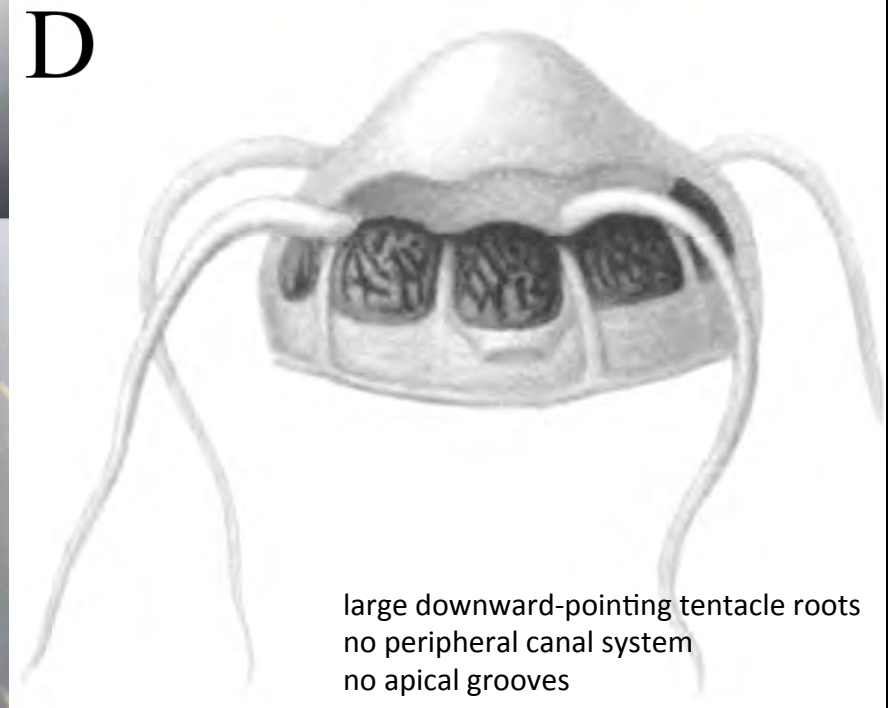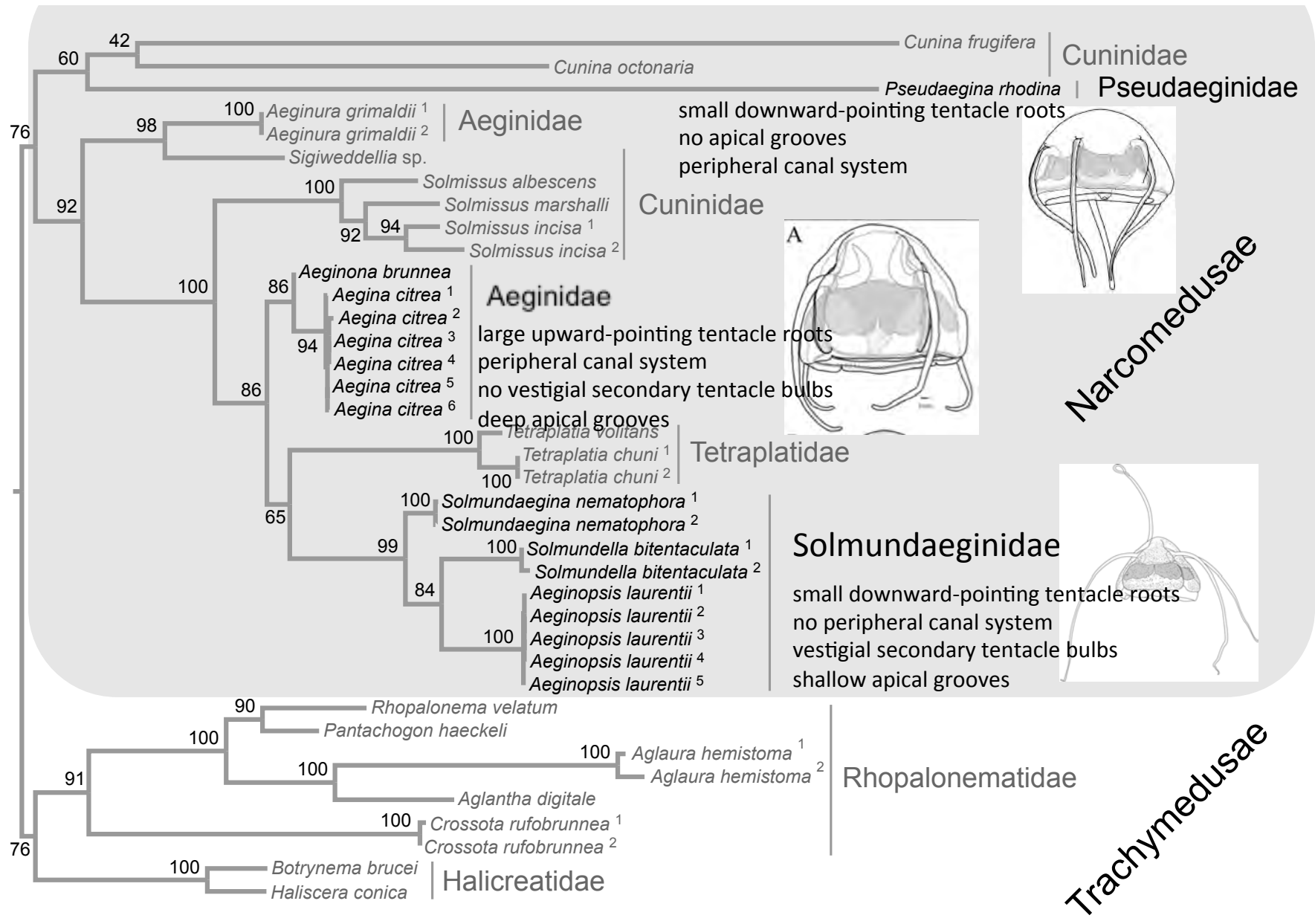
A

B

C

D

large downward-pointing tentacle roots
no peripheral canal system
no apical grooves

Aeginona Lindsay, 2017

Aeginona brunnea (Vanhöffen, 1908)

Cuninidae
Cunina frugifera
Cunina octonaria
Pseudaegina rhodina | Pseudaeginidae

42
60
76

small downward-pointing tentacle roots
no apical grooves
peripheral canal system

100 *Aeginura grimaldii* [1]
Aeginura grimaldii [2]
Sigiweddellia sp.
98
92
Aeginidae

Cuninidae
100 *Solmissus albescens*
Solmissus marshalli
Solmissus incisa [1]
Solmissus incisa [2]
92 94

100

Aeginona brunnea
Aegina citrea [1]
Aegina citrea [2]
Aegina citrea [3]
Aegina citrea [4]
Aegina citrea [5]
Aegina citrea [6]
86
94
86

Aeginidae

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

100 *Tetraplatia volitans*
Tetraplatia chuni [1]
Tetraplatia chuni [2]
100
Tetraplatidae

65

100 *Solmundaegina nematophora* [1]
Solmundaegina nematophora [2]
99
100 *Solmundella bitentaculata* [1]
Solmundella bitentaculata [2]
84
Aeginopsis laurentii [1]
Aeginopsis laurentii [2]
Aeginopsis laurentii [3]
100 *Aeginopsis laurentii* [4]
Aeginopsis laurentii [5]

Solmundaeginidae

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

Narcomedusae

90 *Rhopalonema velatum*
Pantachogon haeckeli
100
100 *Aglaura hemistoma* [1]
Aglaura hemistoma [2]
91
100
Aglantha digitale
100 *Crossota rufobrunnea* [1]
Crossota rufobrunnea [2]
76

Rhopalonematidae

100 *Botrynema brucei*
Haliscera conica
Halicreatidae

Trachymedusae

0.2

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.
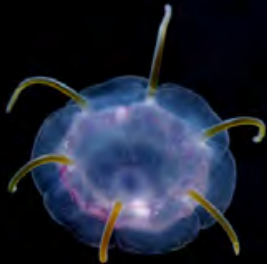
Summary   (Part 1)

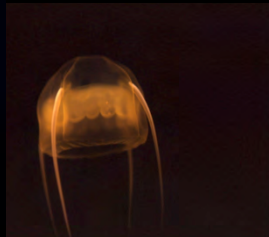- *Aegina citrea* was split into 6 separate species

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.
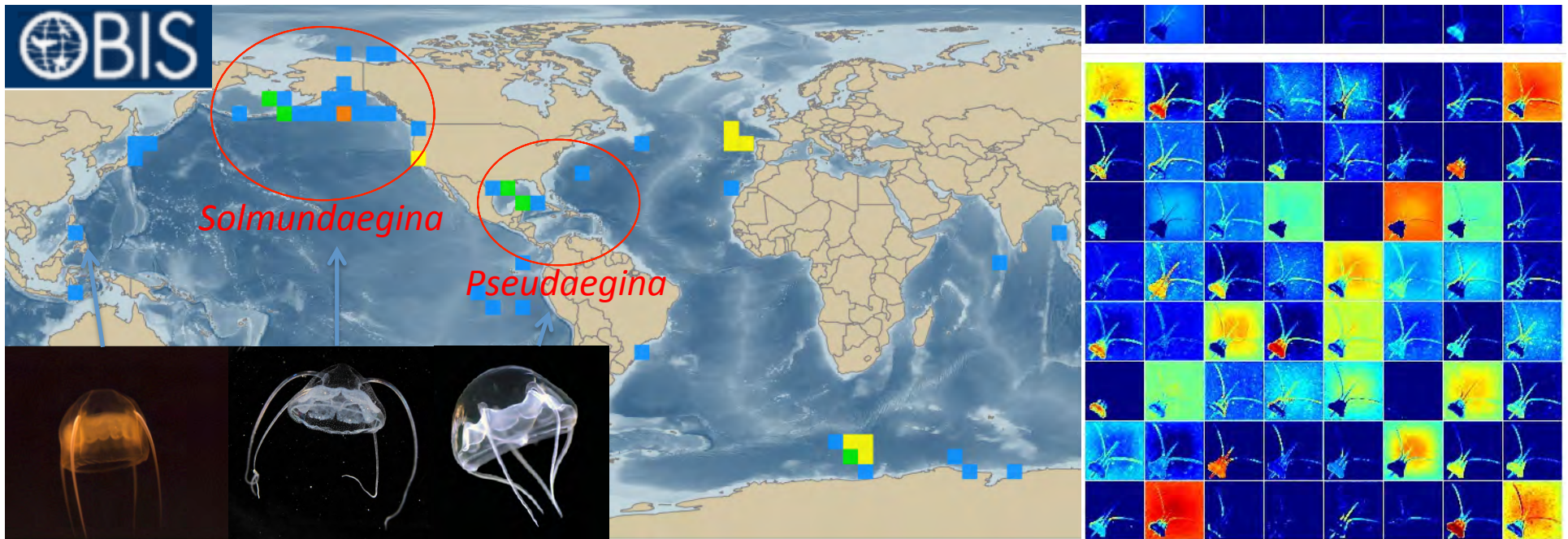
What are their REAL distributions?

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?
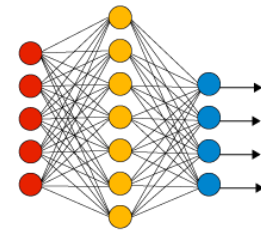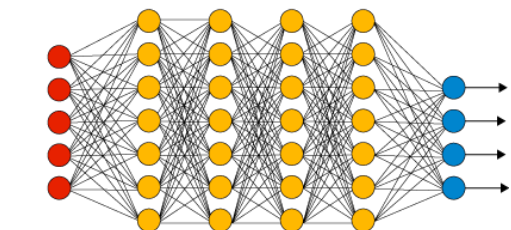
Deep Learning AI

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

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

Image Mean:

```
●●● 端末
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 fi
le 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 fi
le 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 fi
le or directory: '/usr/local/DL-Box/digits-4.0/digits/jobs/20170704-105938-97f4/
status.pickle'

 _____    _____  _____  _____
| |  \/ _|_|  _| | |_| |  \
| |) | | (_| | |  | | | |\ \
|_|__/\_|__|_|_| |_| |_| \_\ / 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 HTT
P/1.1" 304 189 0.003185
```

Explore the db

Choice of network?

GoogleNet [22 layers] gave better results than AlexNet [8 layers] (but still not great..)

70-80% accuracy
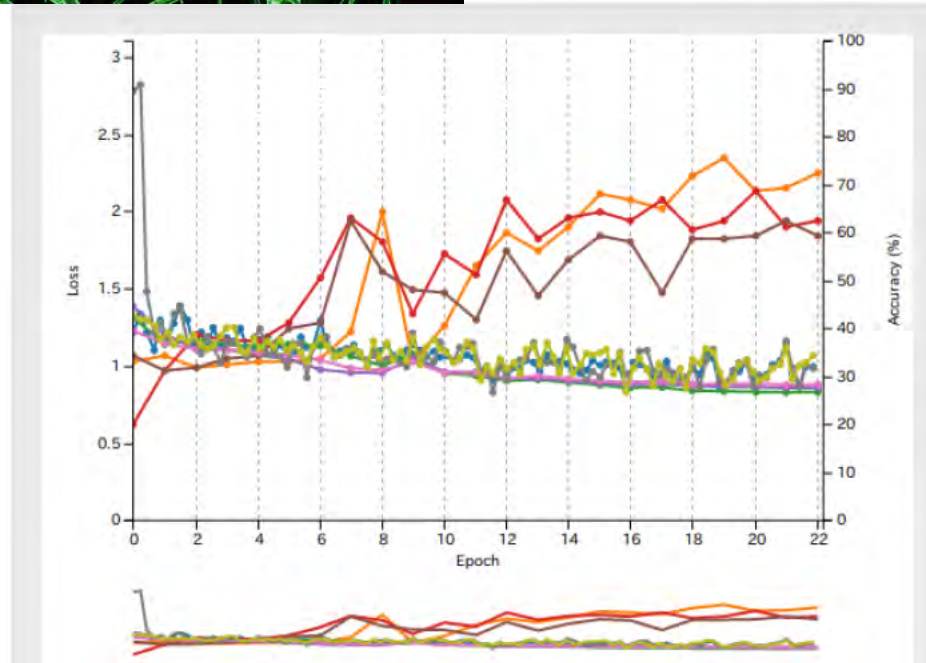
GPU Usage

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
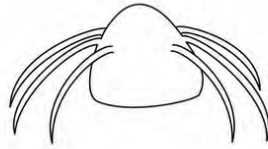
Predictions

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

Description	Statistics	Visualization
data Activation	Data shape: [1 224 224] Mean: -1.67496 Std deviation: 33.4134	
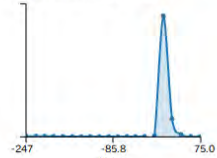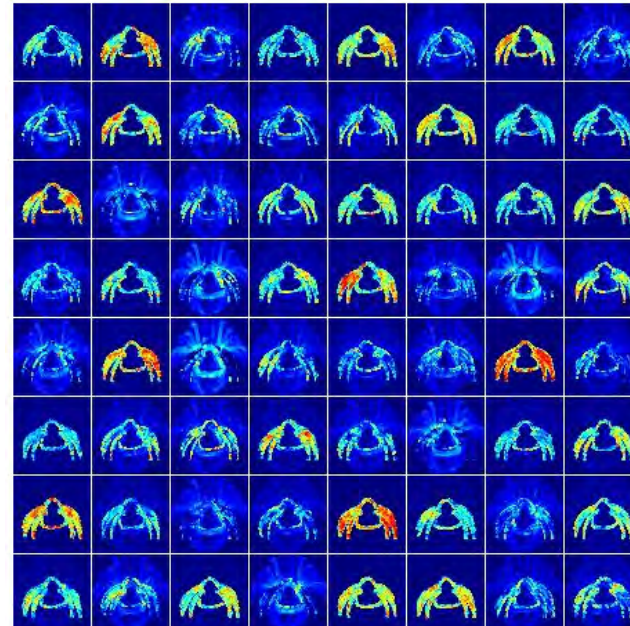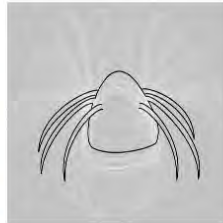
tentacle integrity only up to second convolution

Not-so-Deep Learning better?
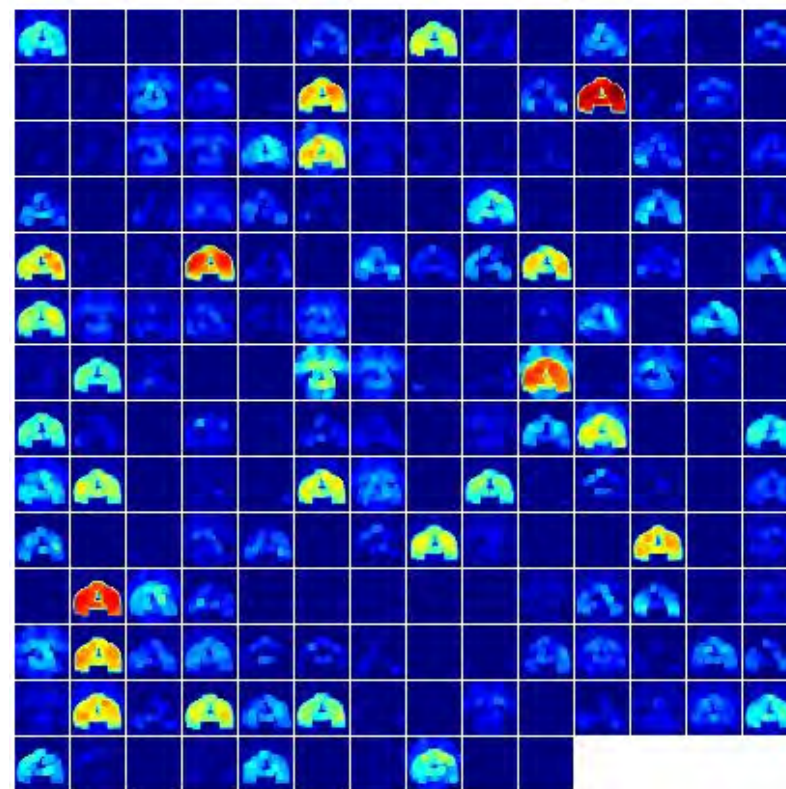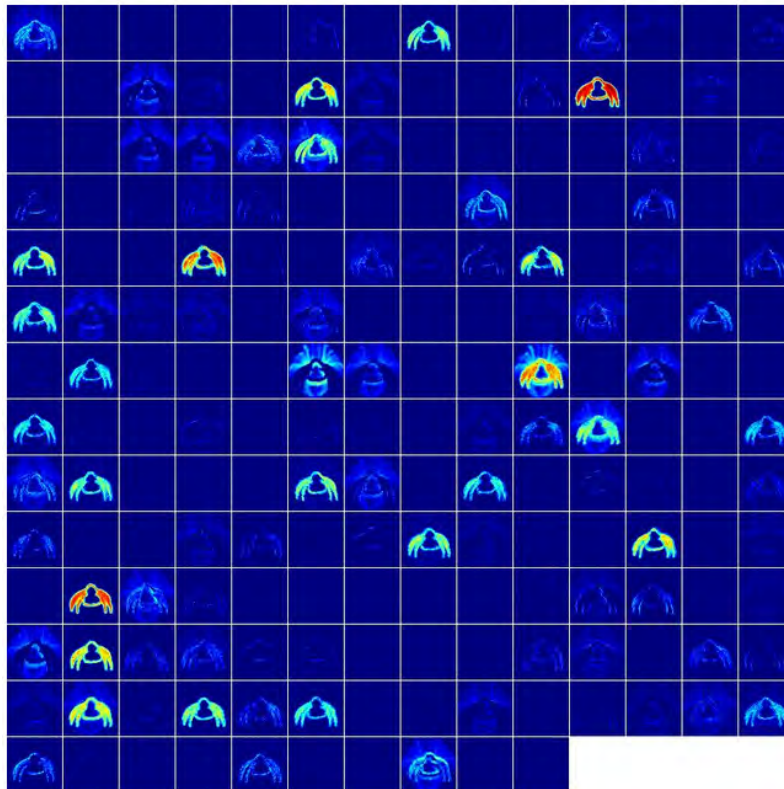
honbantestgoogle1 Image Classification Model

59% correct

Predictions

DeepLearning Solmundella bitentaculata	59.34%
DeepLearning Aegina citrea	24.16%
DeepLearning Aegina rosea	16.49%
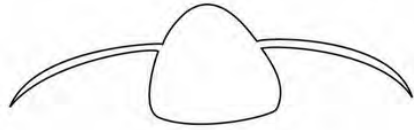
Description	Statistics	Visualization
data Activation	**Data shape:** [1 224 224] **Mean:** 1.75593 **Std deviation:** 24.6611	
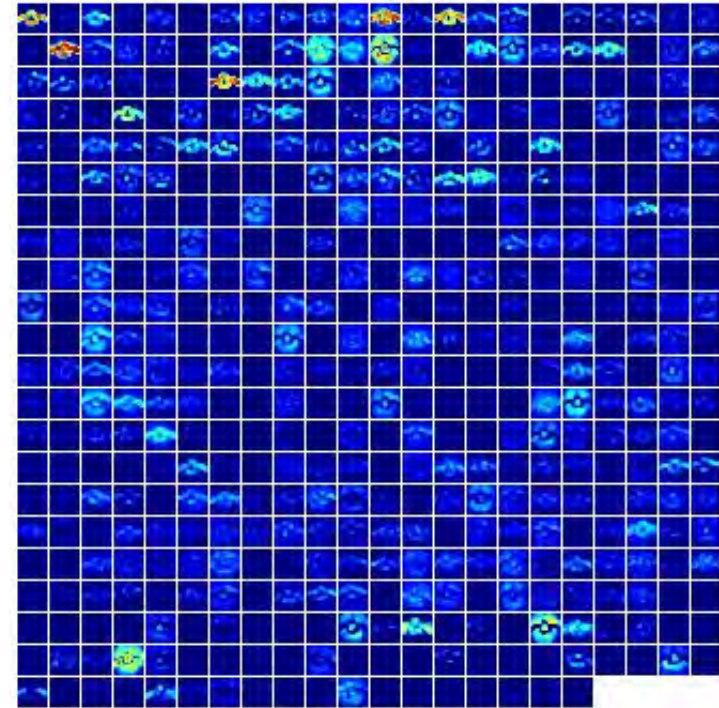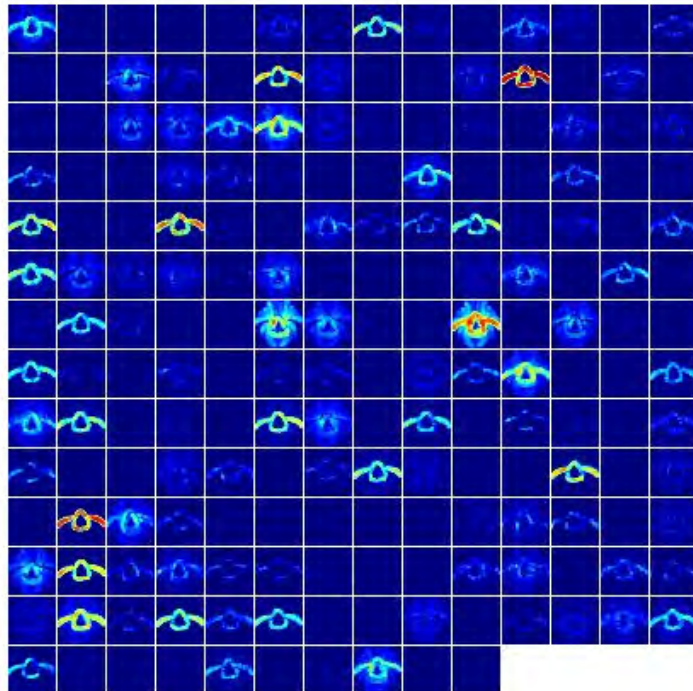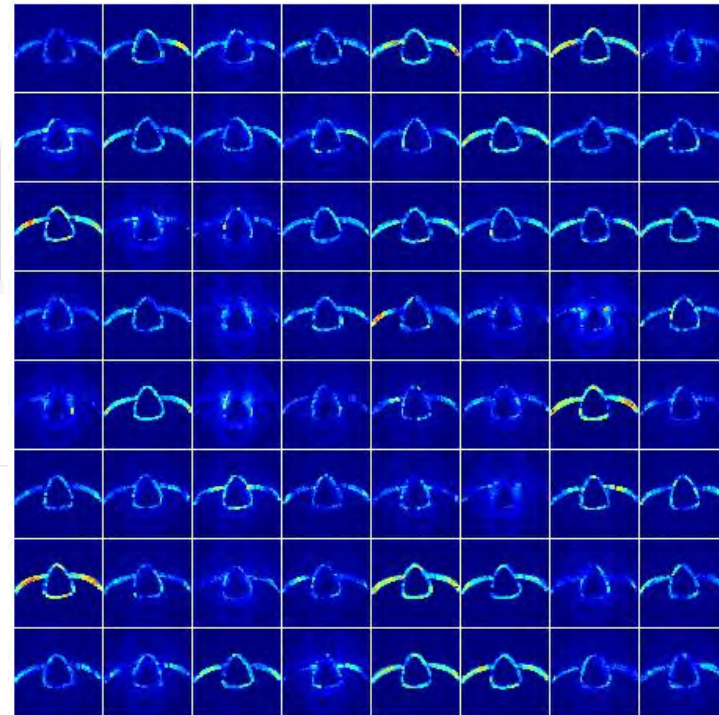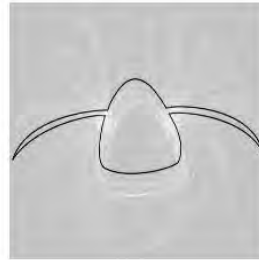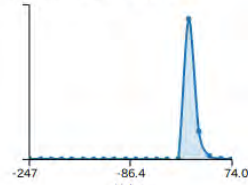
Made an in situ image training set for 2-6 tentacle Narcomedusae using framegrabs from JAMSTEC ROV & HOV videos

convert videos to still frame series, extract jellies manually with proprietary software,
flip and rotate images to increase number of images in training set

500-700 images per species

for humans line drawings look more like jellyfish

for Deep Learning this worked better!

Image Mean:

Explore the db

Data shape
Mean: 30.4
Std deviati
0.00

after 22 training iterations still low accuracy

only 11 training iterations for 100% accuracy
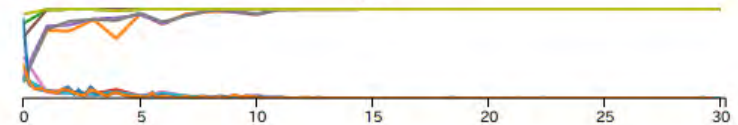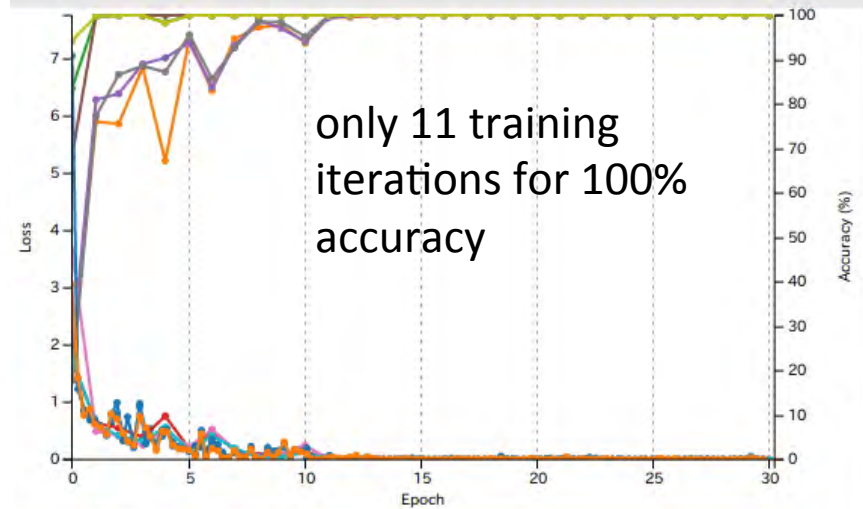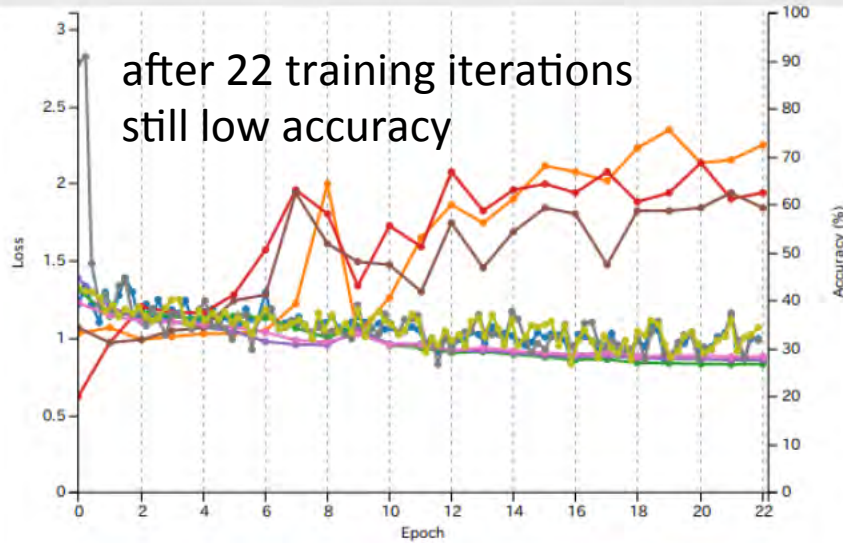
Tried with "real" *Aegina citrea* framegrabs.
Classified image from internet with 99.9% accuracy!!

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

Aeginura grimaldii also 99.9% accuracy!!

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

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

Need to ensure training set contains all species!!!

Solmundaegina nematophora was correctly identifed 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?

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.
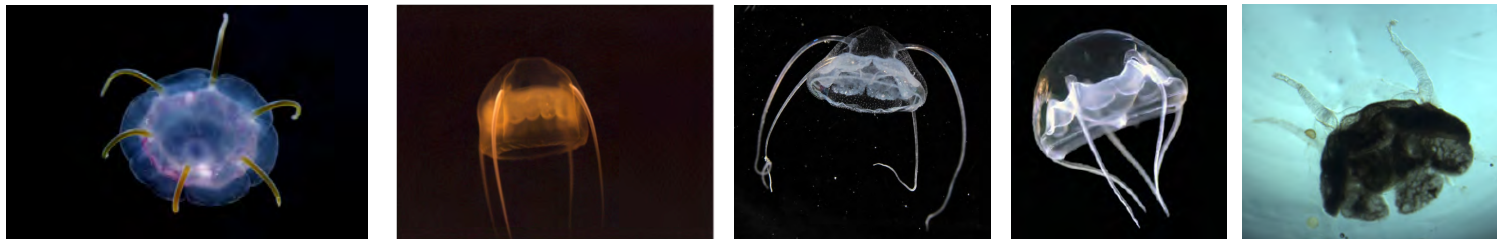
（caveat: only tested with Narcomedusae so far）
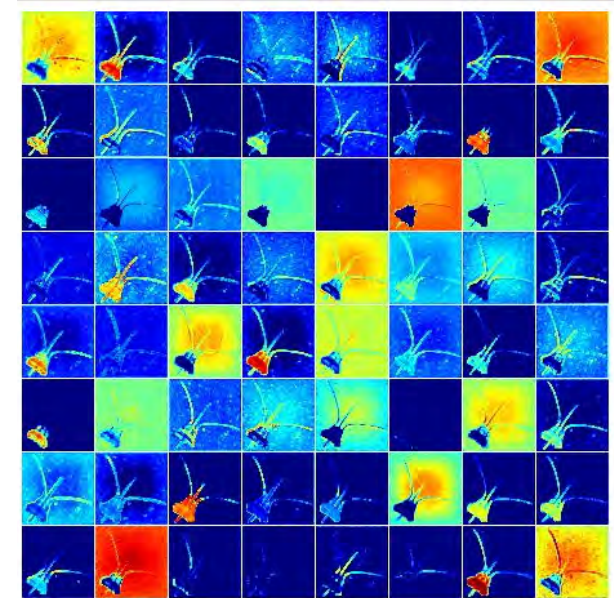
accurate biogeographic data may be in our sights!

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.
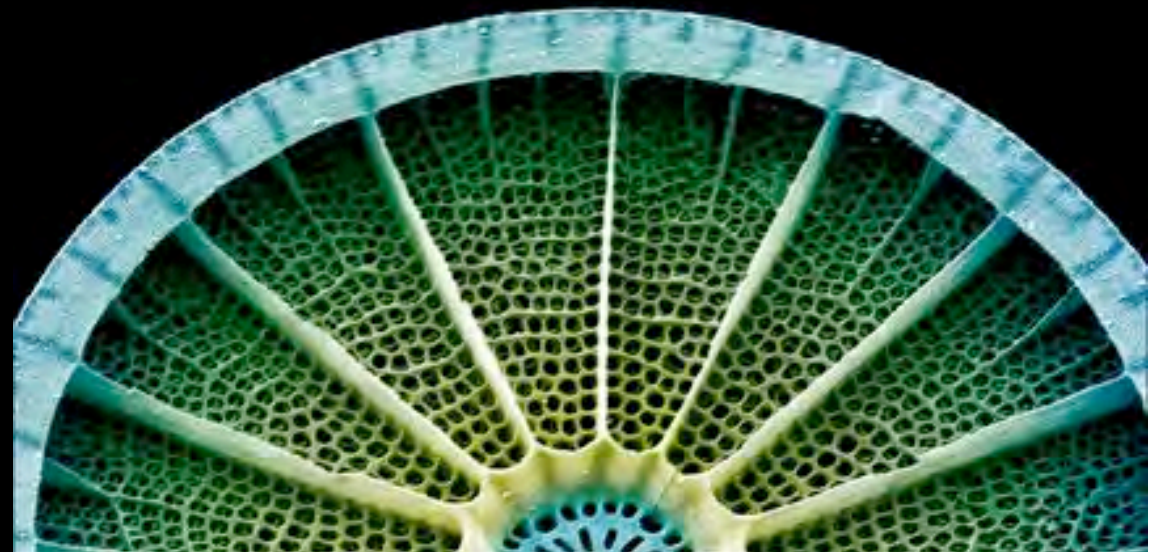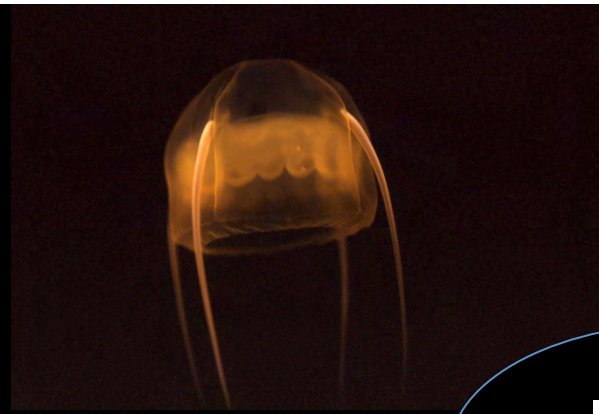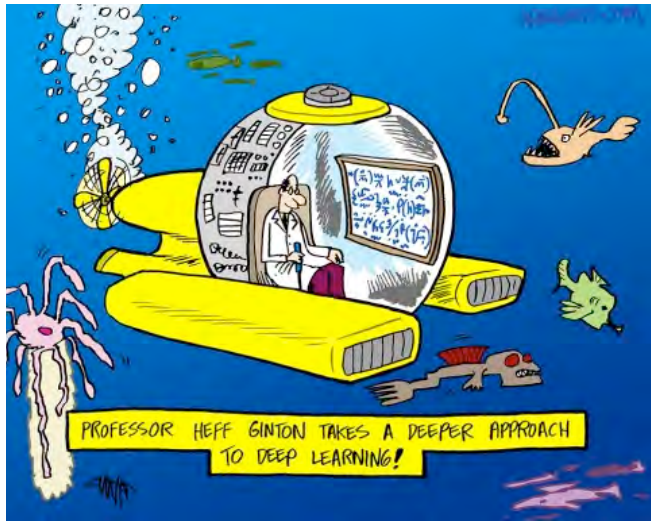
OXFORD

MARINE PLANKTON

A Practical Guide to Ecology, Methodology, and Taxonomy
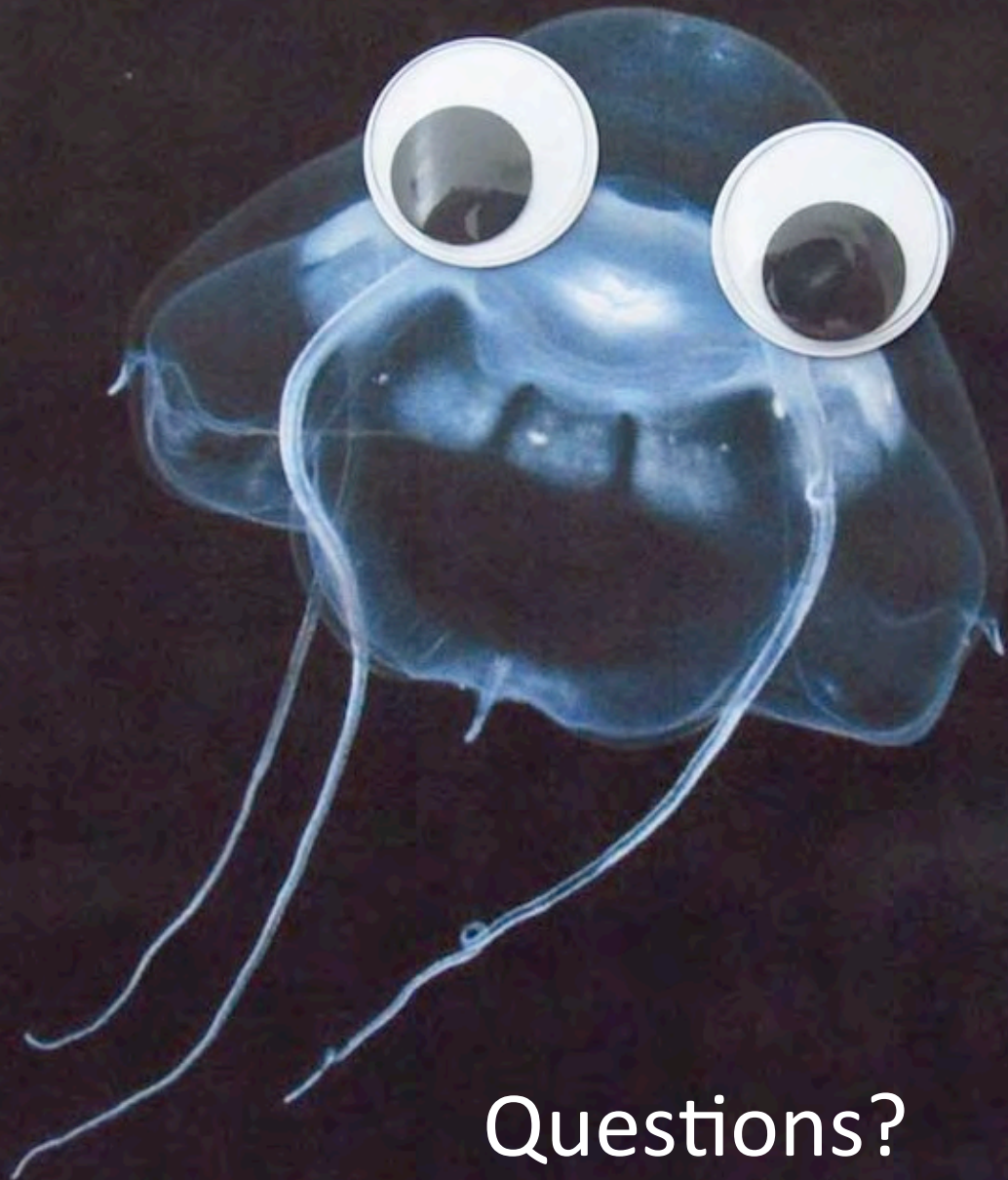
Edited by
Claudia Castellani & Martin Edwards

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?