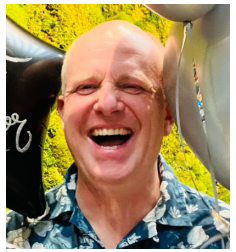
Species Classification Automation for Microfossil Photomicrograph Images (Scampi)

David Wade*, Sissa Stefanowicz, Alex Cullum & Erik Anthonissen






Palynology Slide Analysis

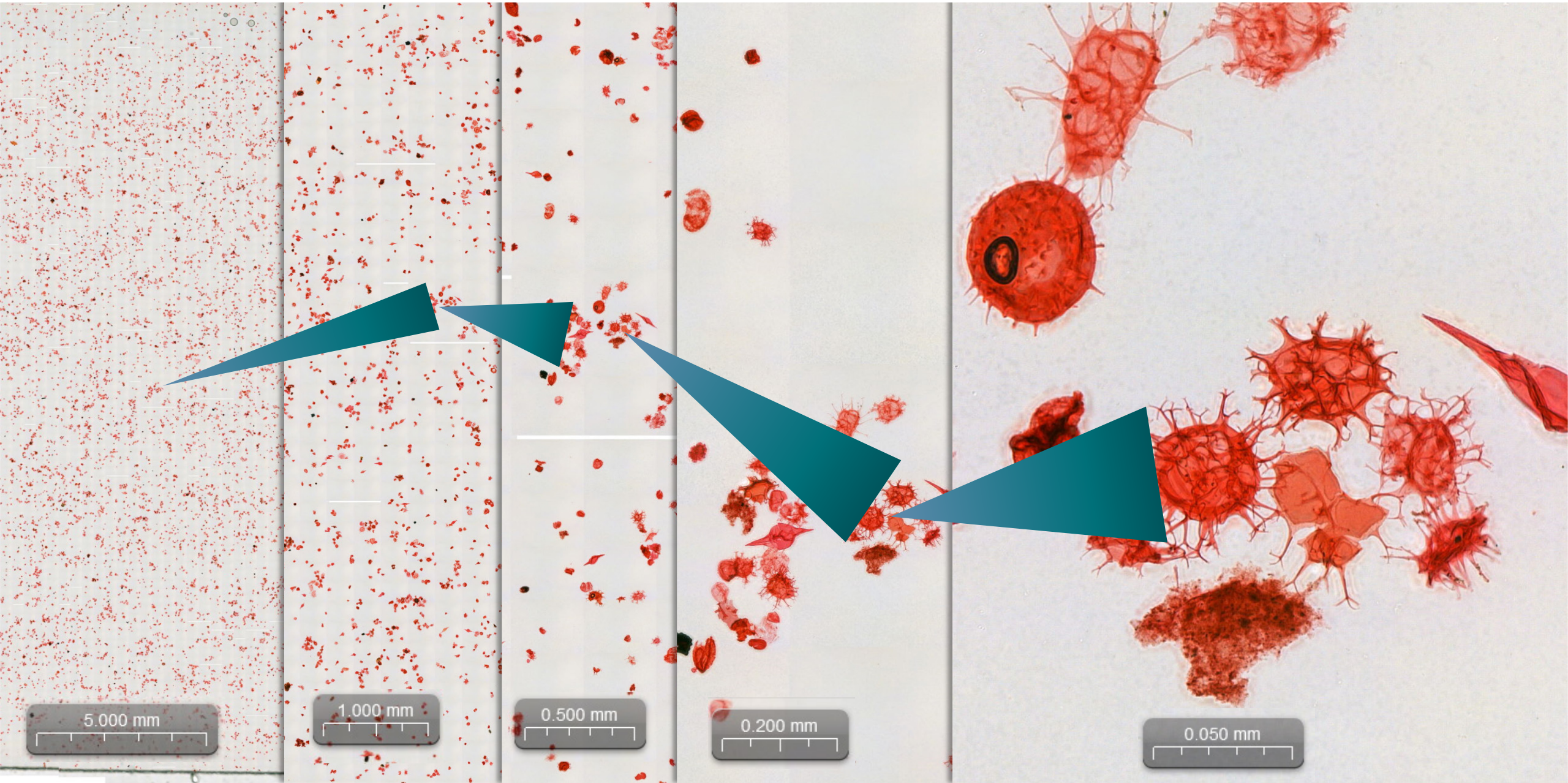
Traditional approach:

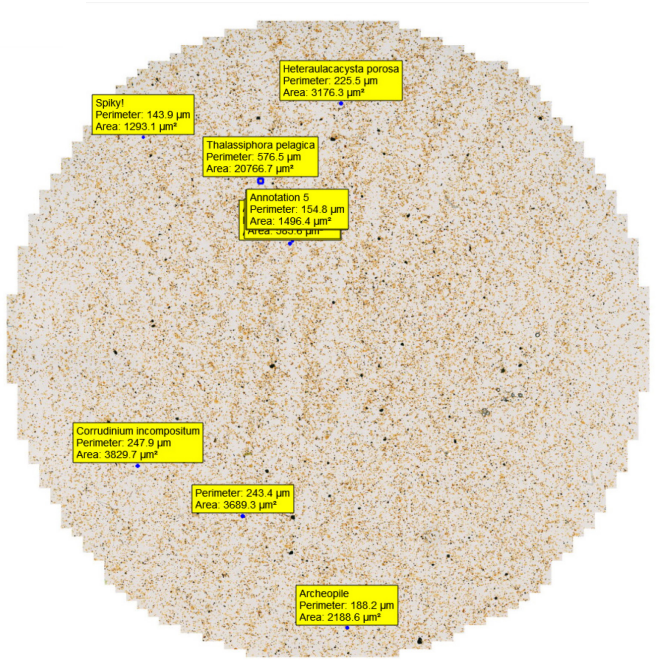
- Long hours with a microscope
- Make species counts
- Few practitioners

equinor

NORWEGIAN OFFSHORE DIRECTORATE

05 March 2024

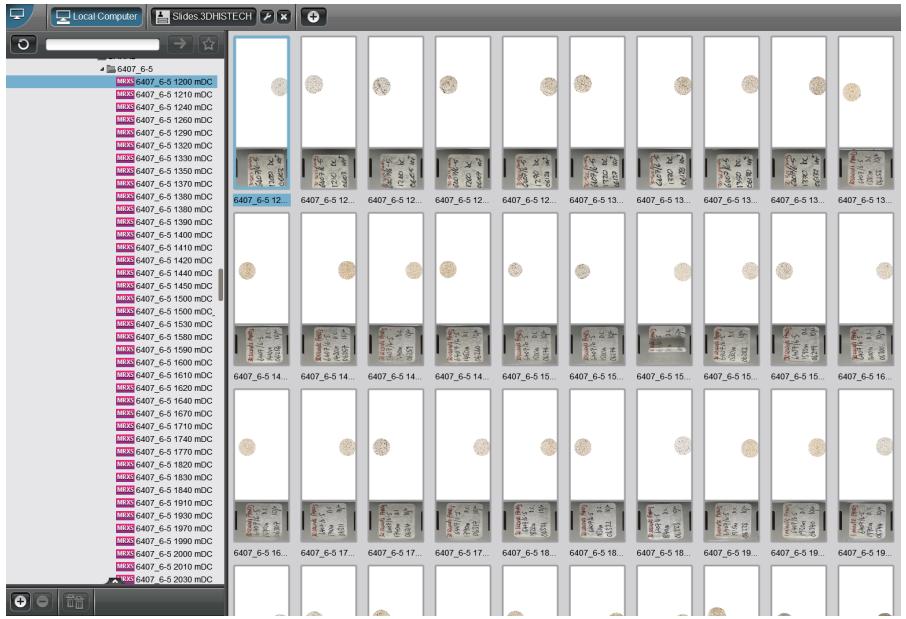

Zoom : Whole slide
... to ...
Individual fossils
5.000 mm
1.000 mm
0.500 mm
0.200 mm
0.050 mm



6
Gigapixels

=

~25,000
Palynomorphs



x80
slides per well

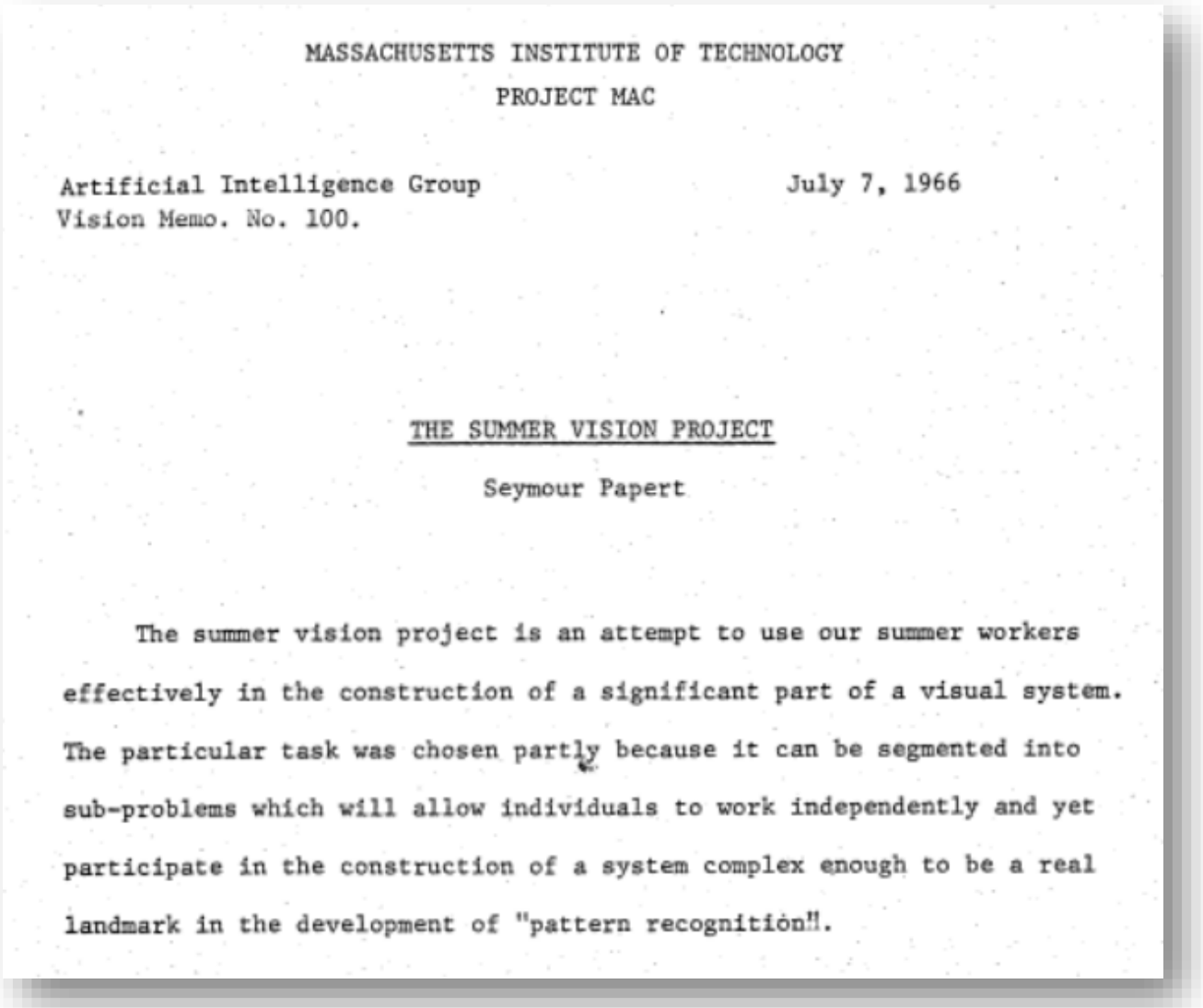
x1800
wells

=

~3 billion
Palynomorphs
> 109 million
200TB

Computer-Vision Origins

- **1966** – ah, we should be able to crack this over the summer…



MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group — July 7, 1966
Vision Memo. No. 100.

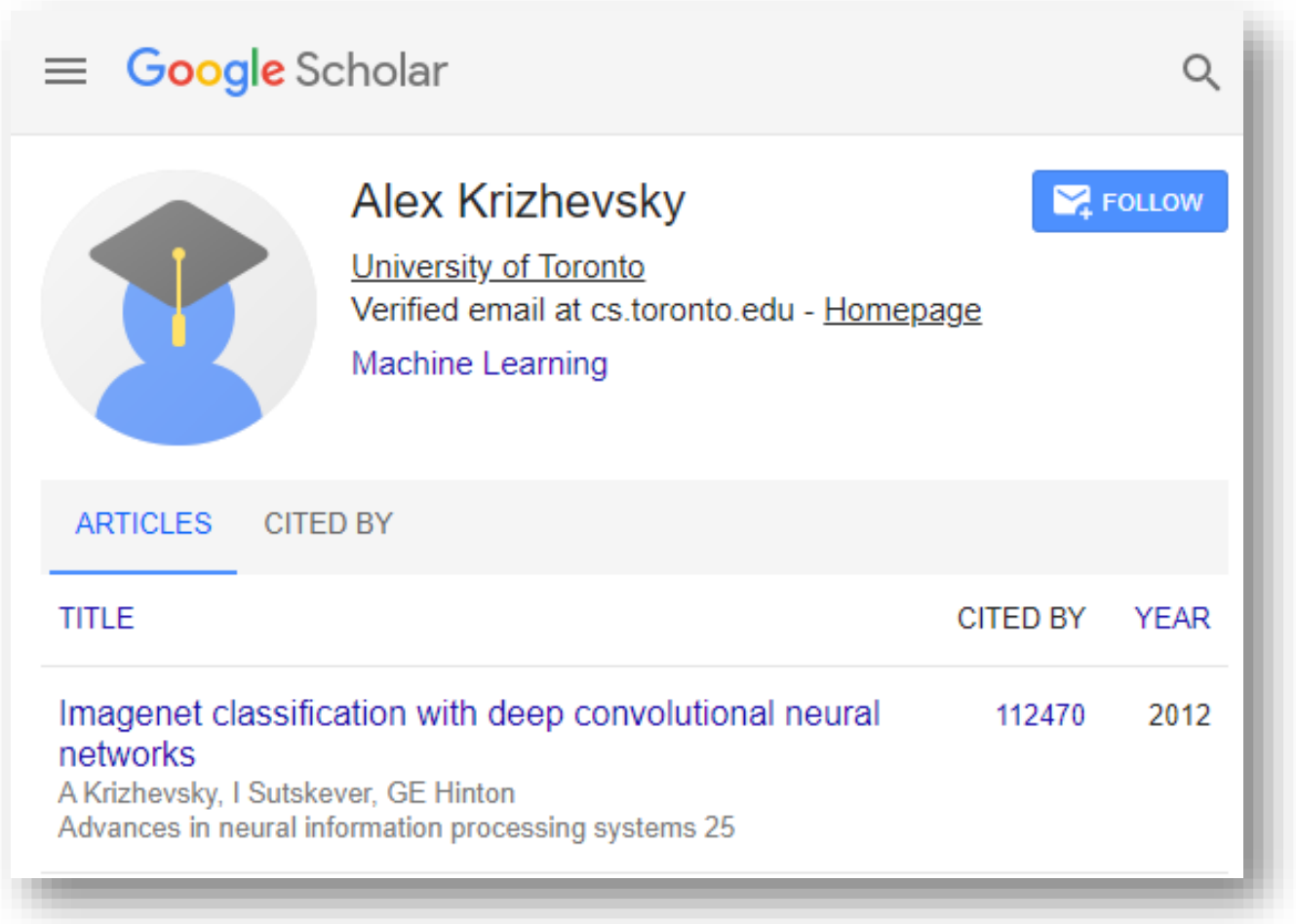
THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

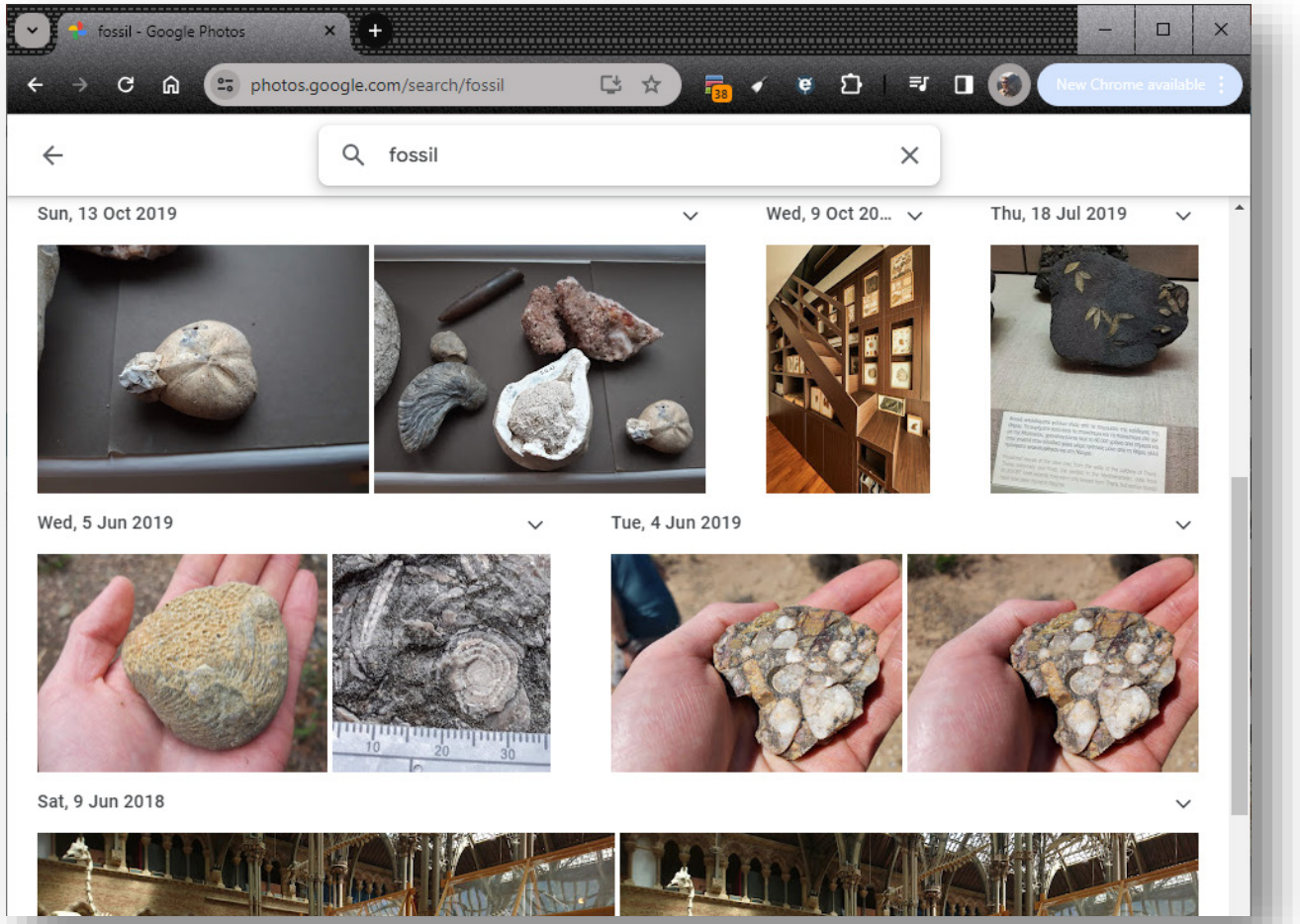
Computer-Vision Origins

- **1966** – ah, we should be able to crack this over the summer…
- **2012** – major breakthrough using DNNs to classify images



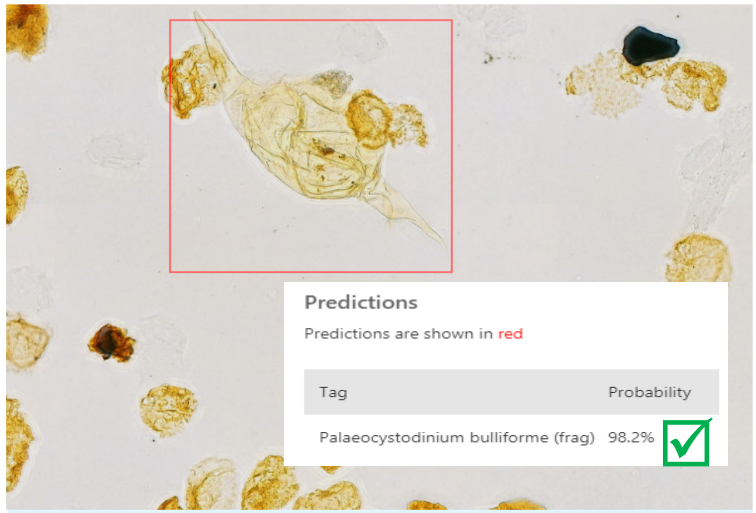
Computer-Vision Origins

- **1966** – ah, we should be able to crack this over the summer…
- **2012** – major breakthrough using DNNs to classify images
- **Now** – readily available commodity



Example results from supervised learning

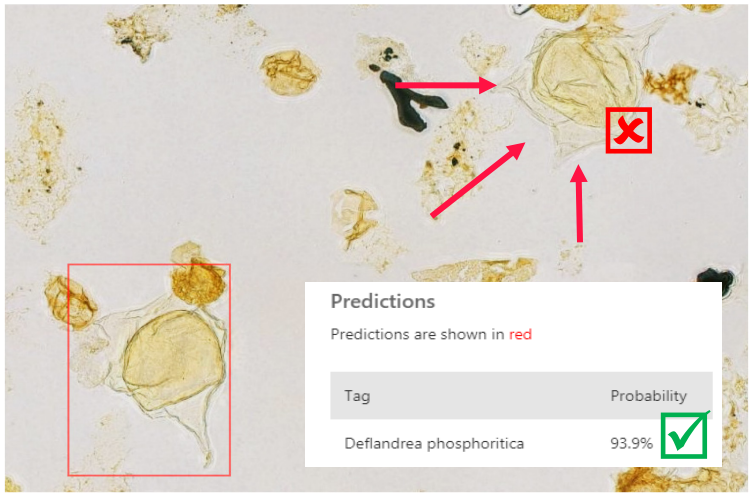
The Good



Correct prediction on image from outside training set

- ✓ *Palaeocystodinium* is here!
- ✓ No false positives

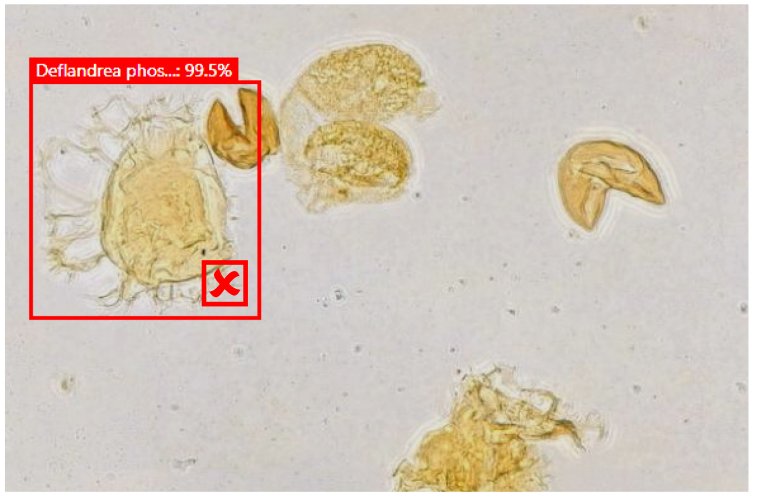
The Bad



Incomplete prediction on image from outside training set

- ✓ *Deflandrea phosphoritica here*
- ✓ Second *Deflandrea* missing?

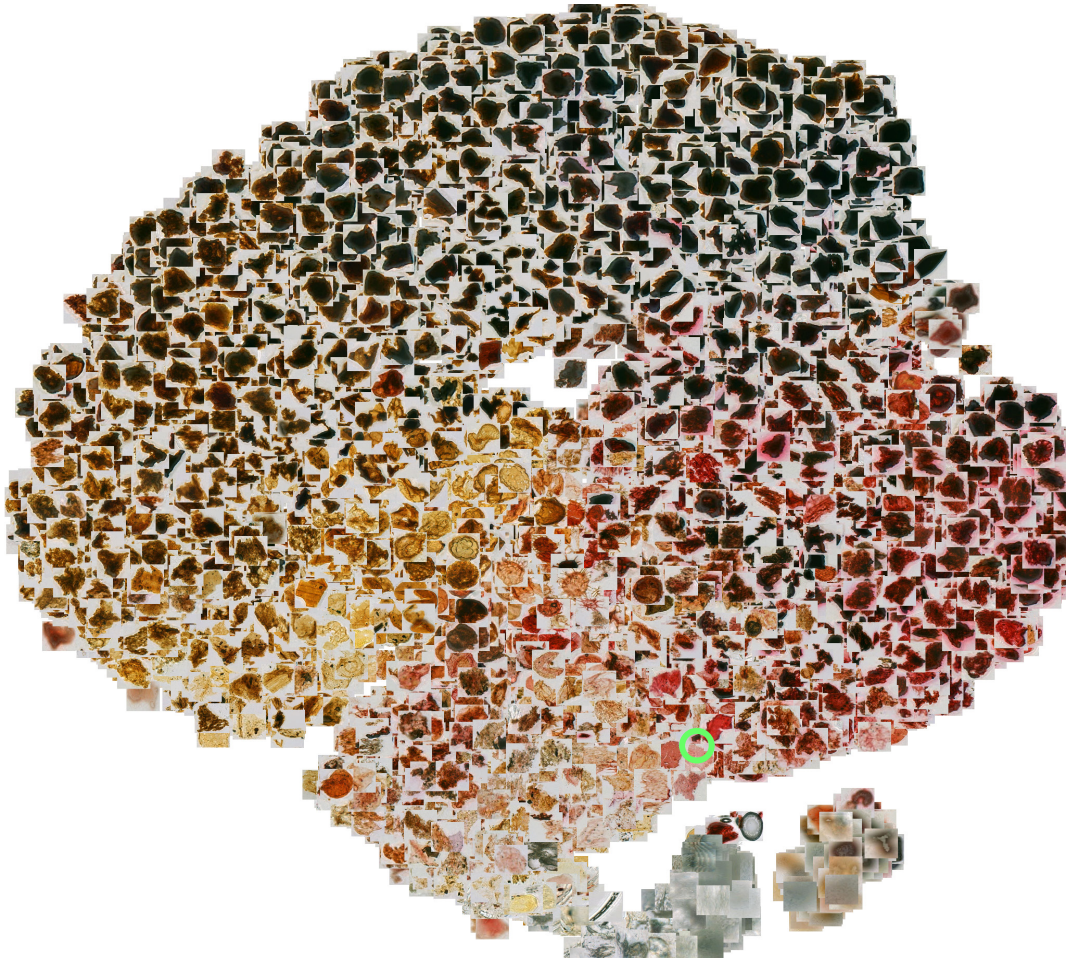
The Ugly

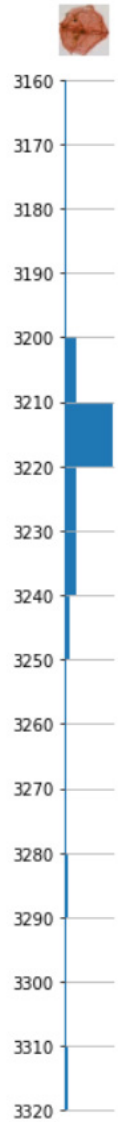


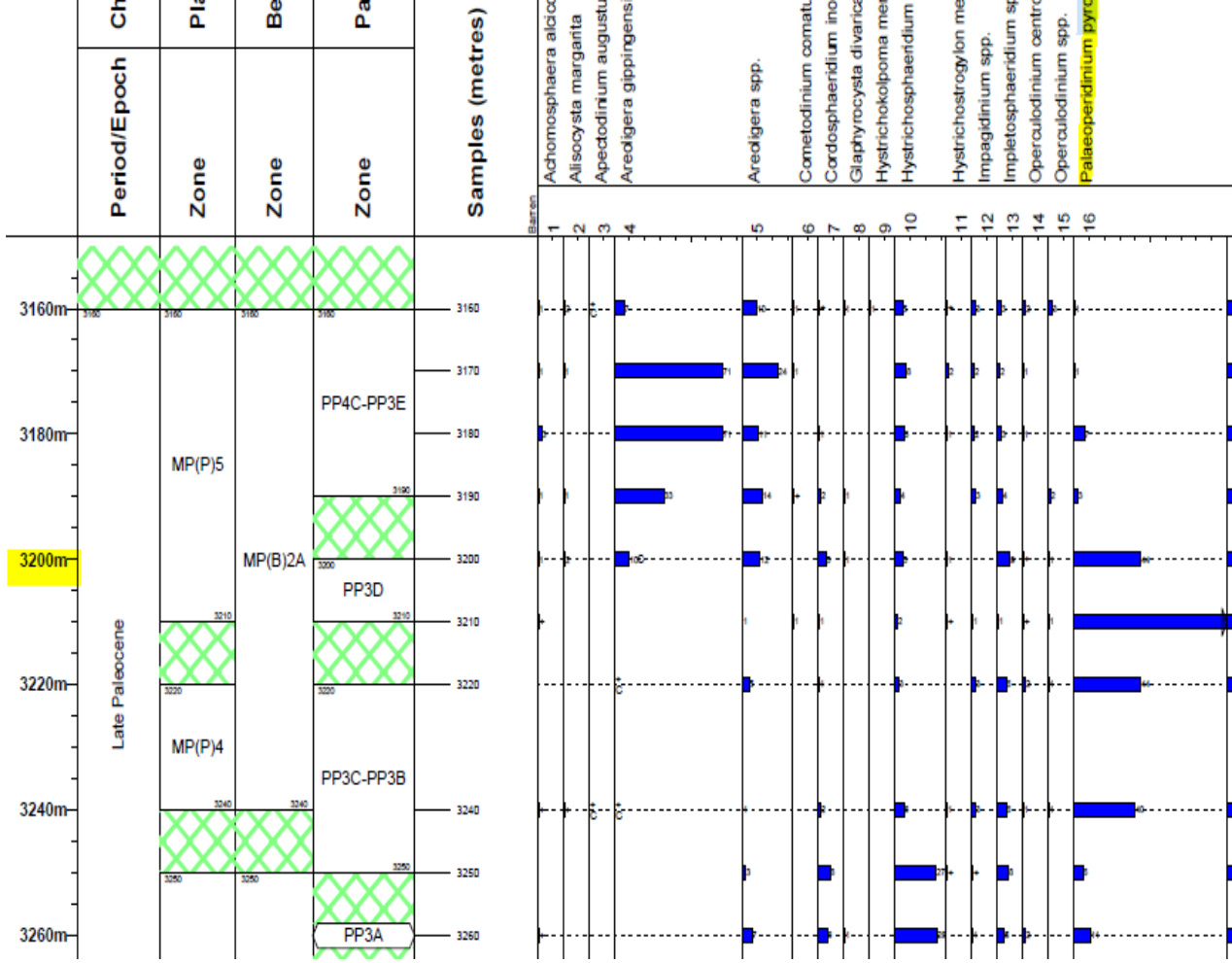
Confident and wrong

- ✓ *Not a Deflandrea*
- ✓ Ignored pollen

t-SNE plot of image embeddings

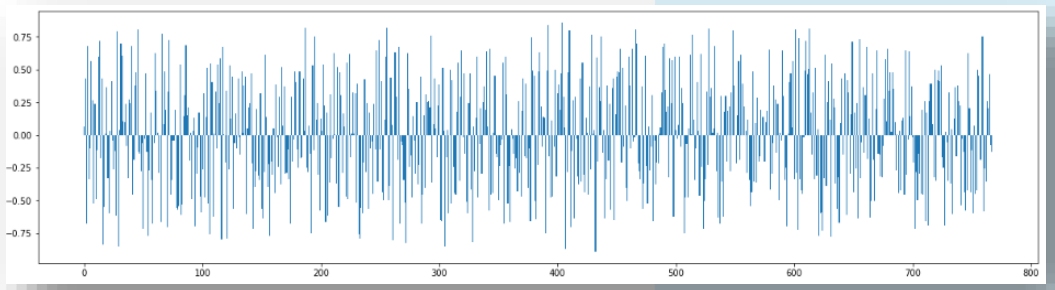




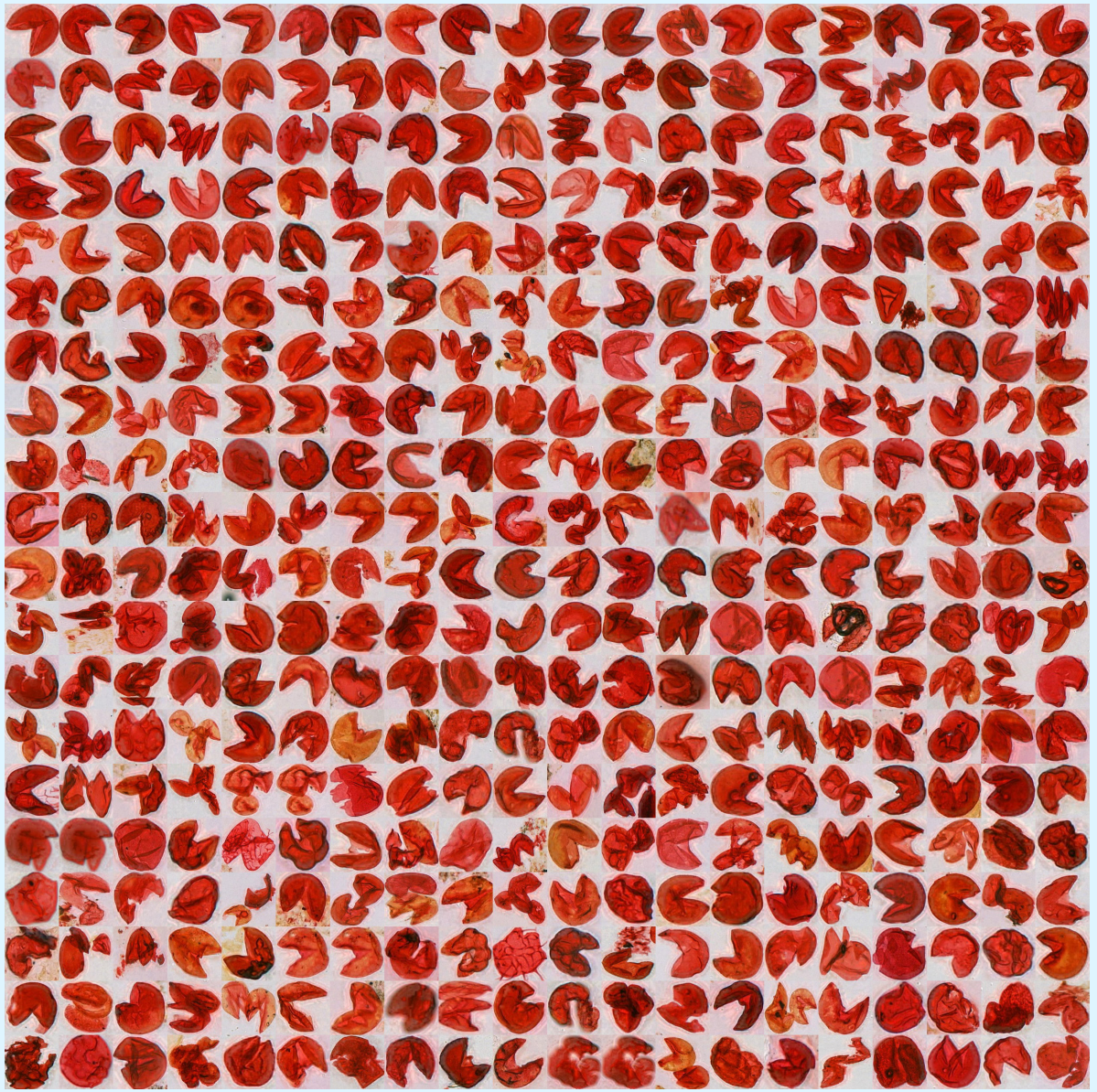


Content-Based Image Retrieval

Query :

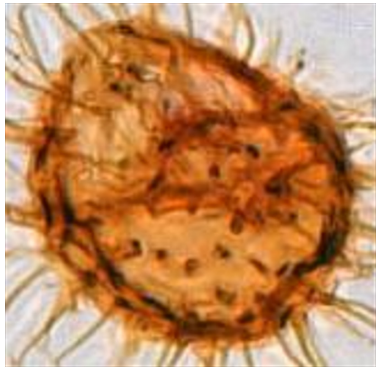


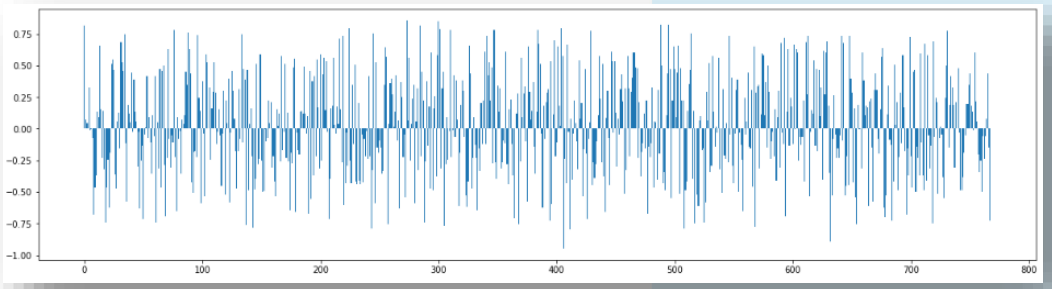
Result :



Content-Based Image Retrieval

Query :

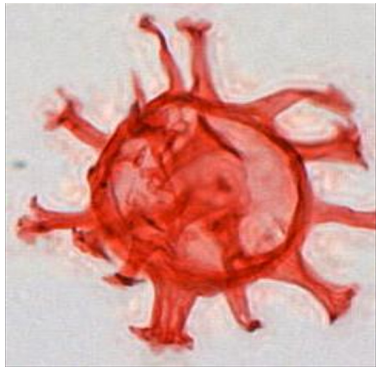


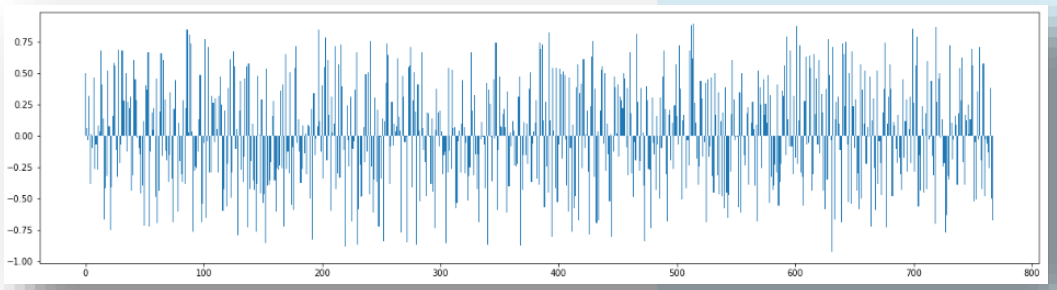


Result :

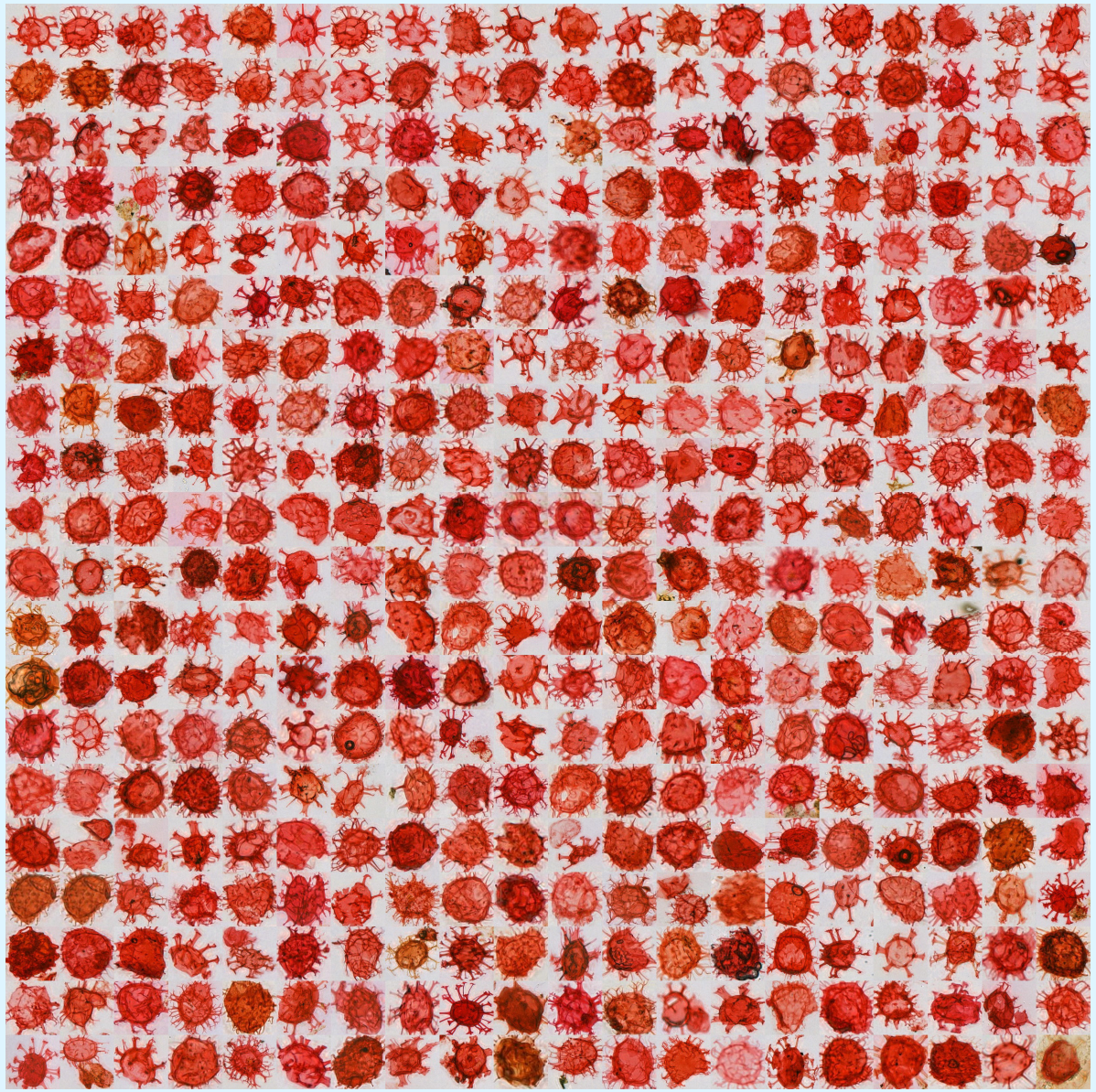
Content-Based Image Retrieval

Query :



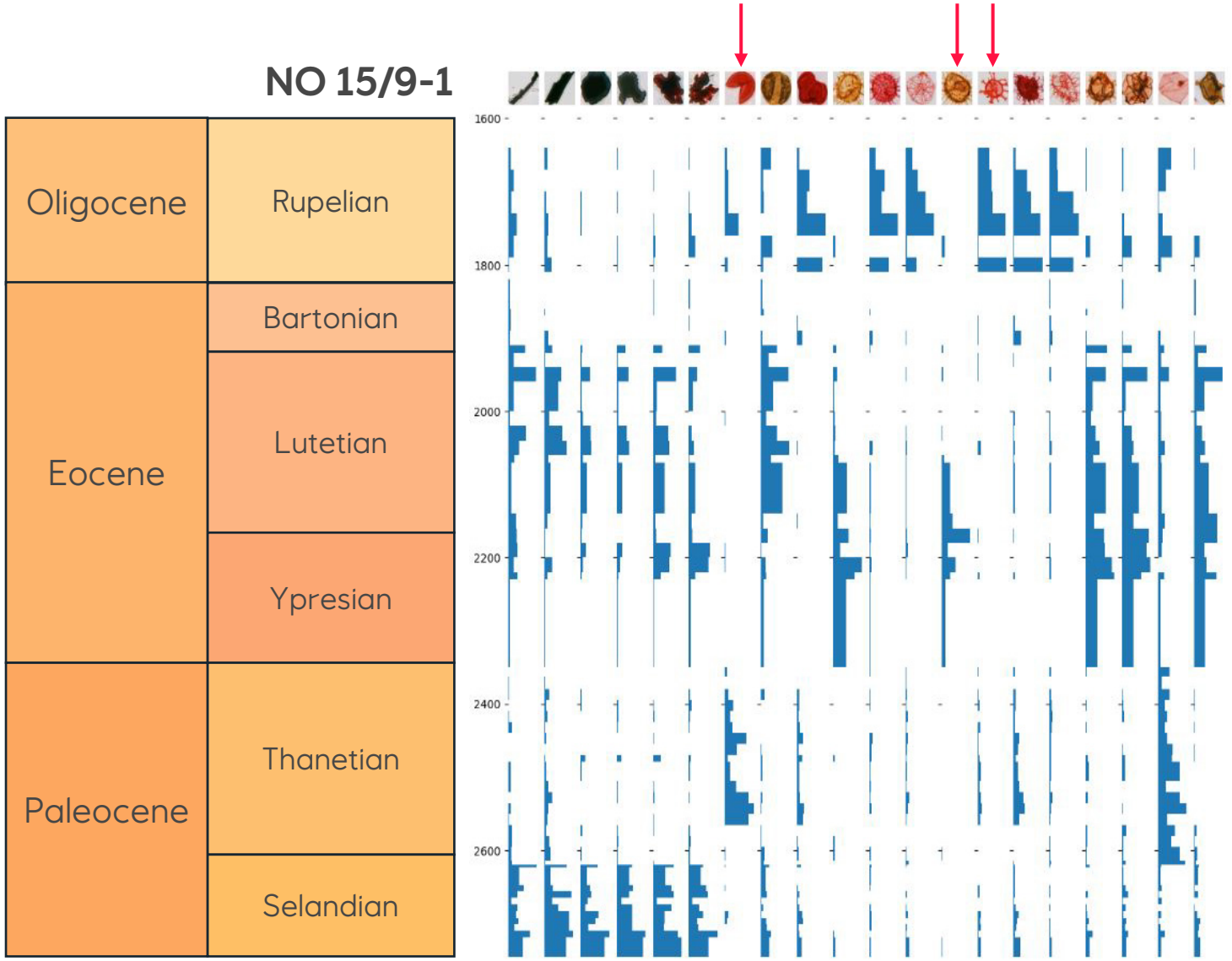


Result :



Chronostratigraphy

- Define multiple search crops
- Find depths of matching crops
- Plot histograms of matchs
- Compare with well picks



Use Case : Formation Zonation

Iteration #1 : Queries from not-a-palynologist

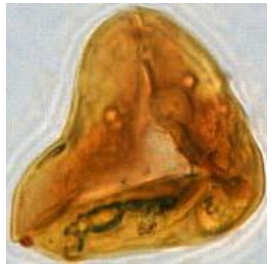
This works surprisingly well:

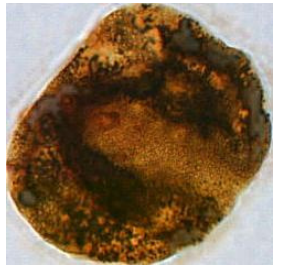
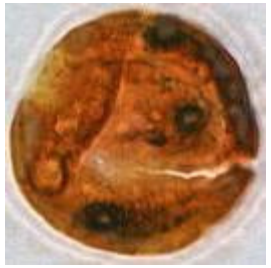
… finds things which look like the queries

… but what do the results mean?!

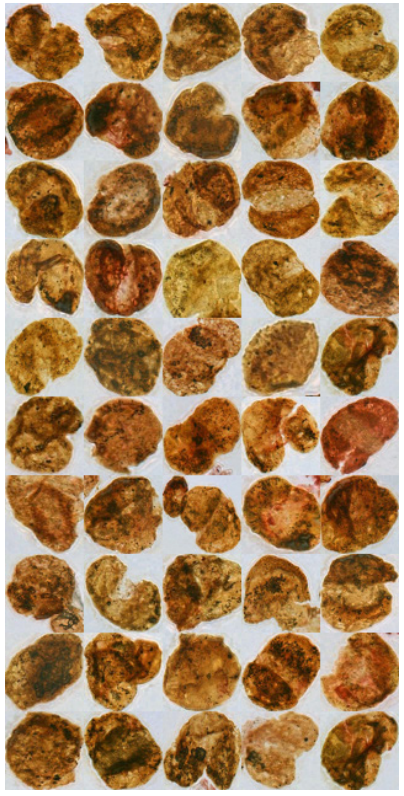
Well A – query results





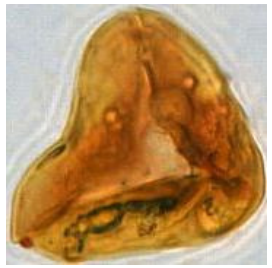



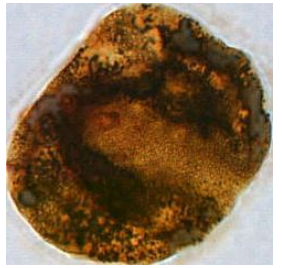
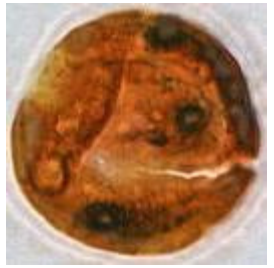


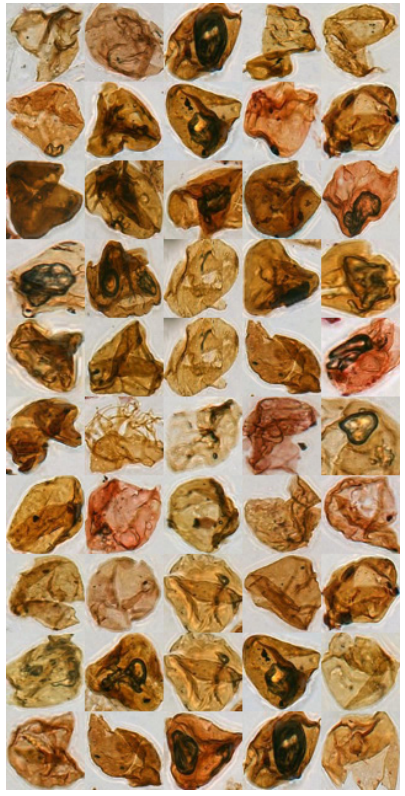



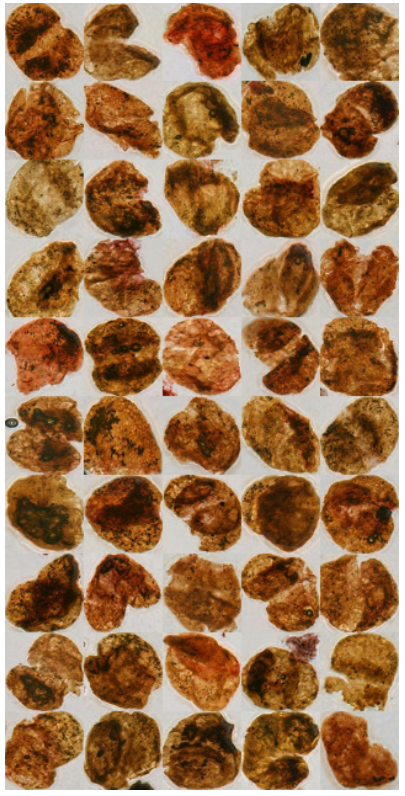
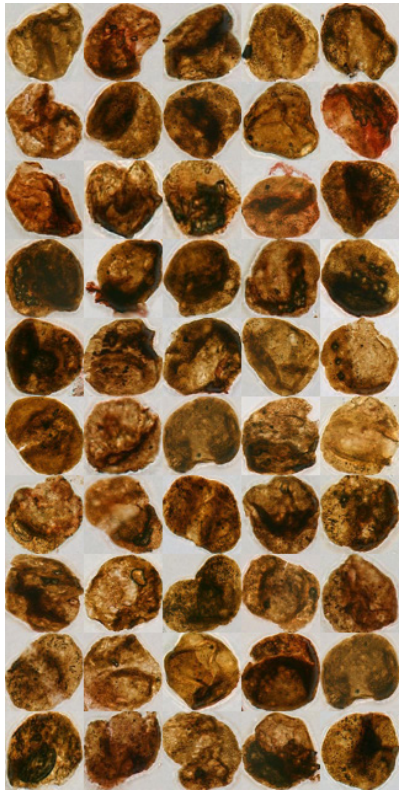
Well B – query results





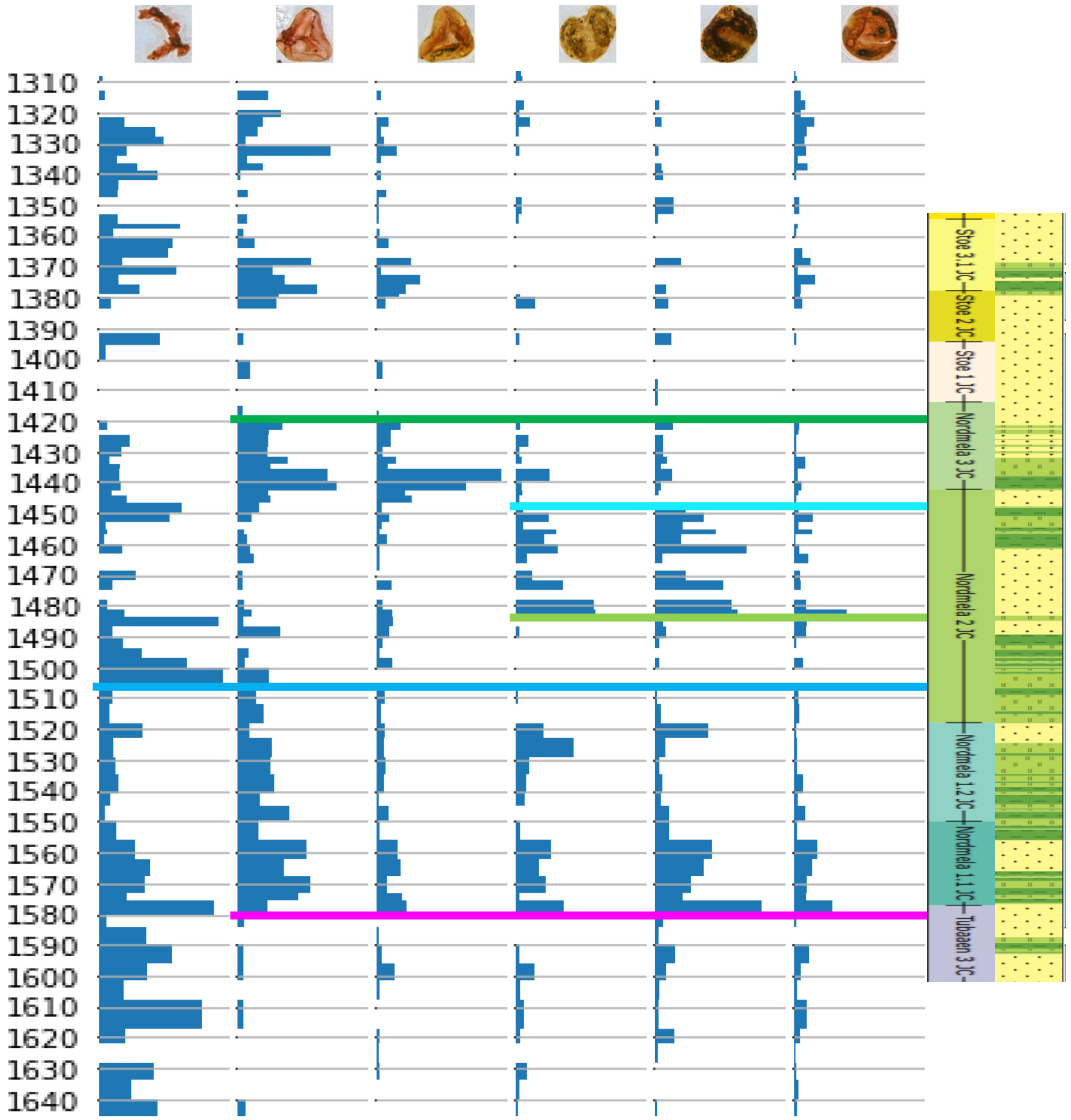



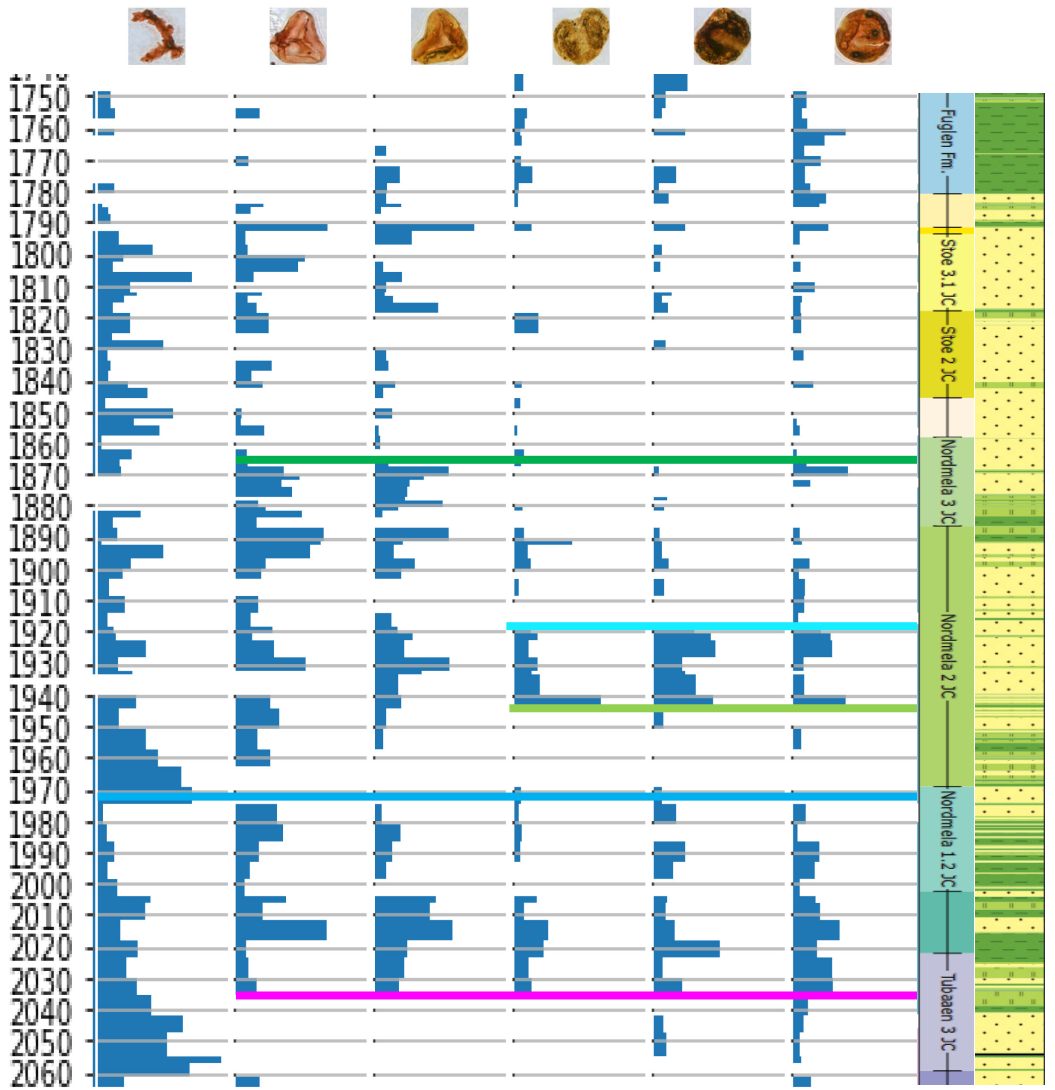

Scampi charts

Well A

Data scientist finds some interesting images

Well B





Use Case : Formation Zonation

Iteration #2 : One real palynologist runs queries

Find *exactly* what we're looking for:

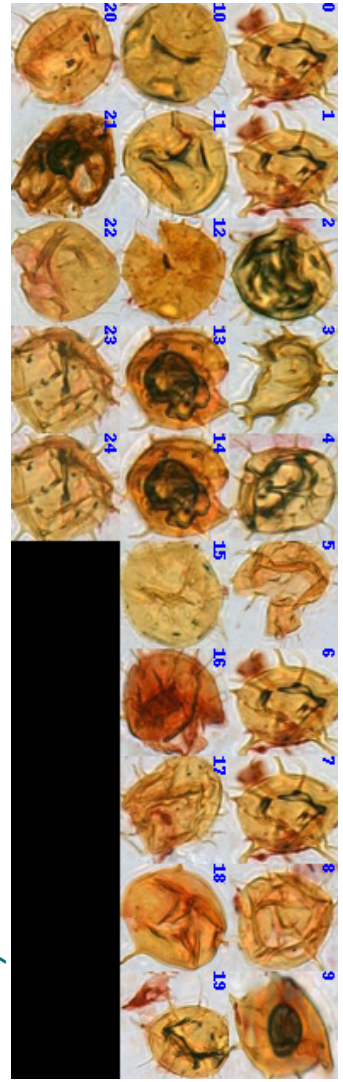
… interpretable results, fast!

… consistent patterns! (which *nearly* agree with official Equinor zonation)

… useful information for the asset!

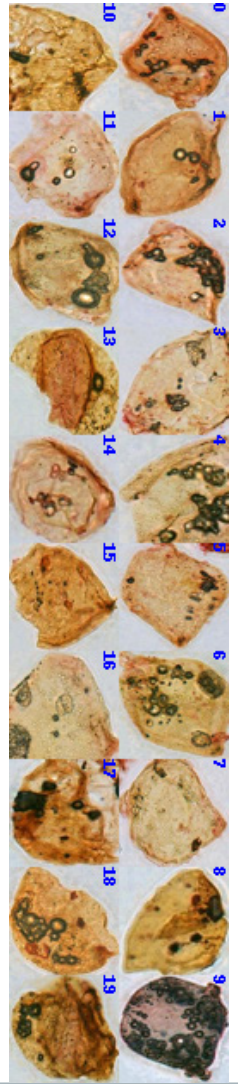
Well A – query results

Micrhystridium

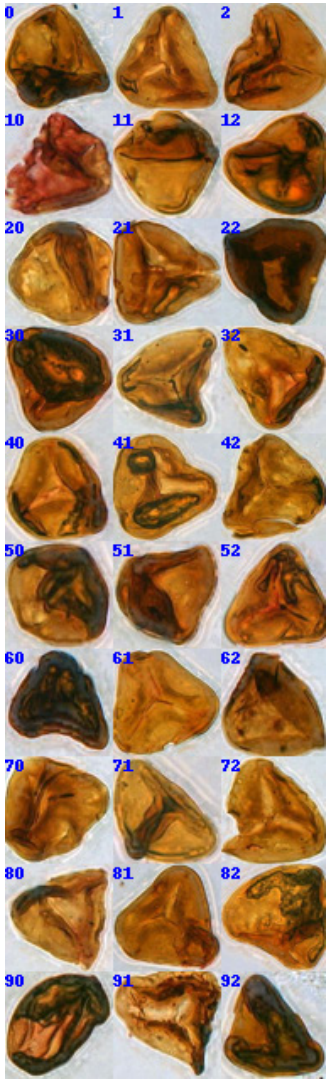


Limbicysta bjaerkei

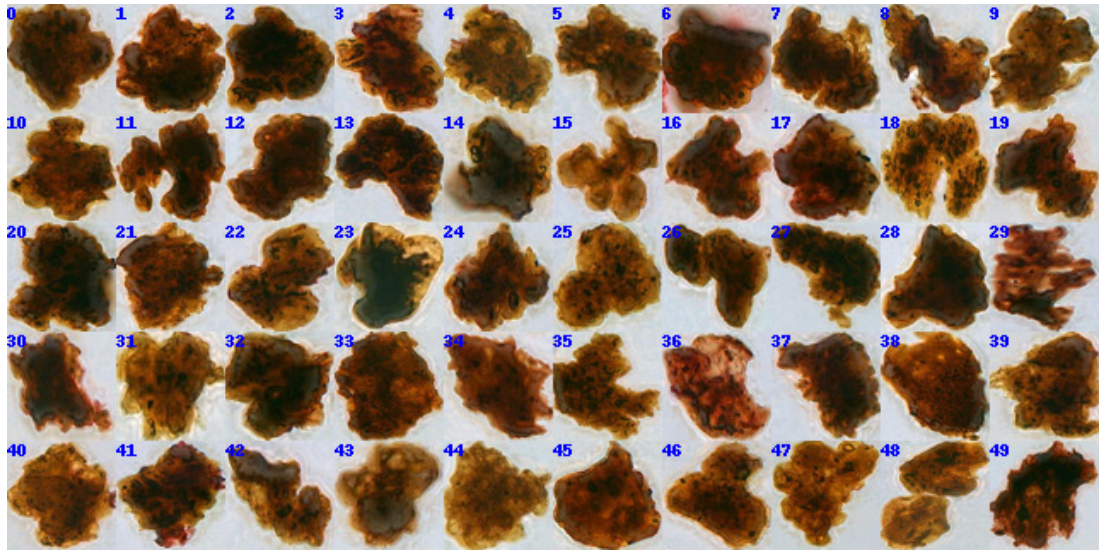
Nannoceratopsis spp.



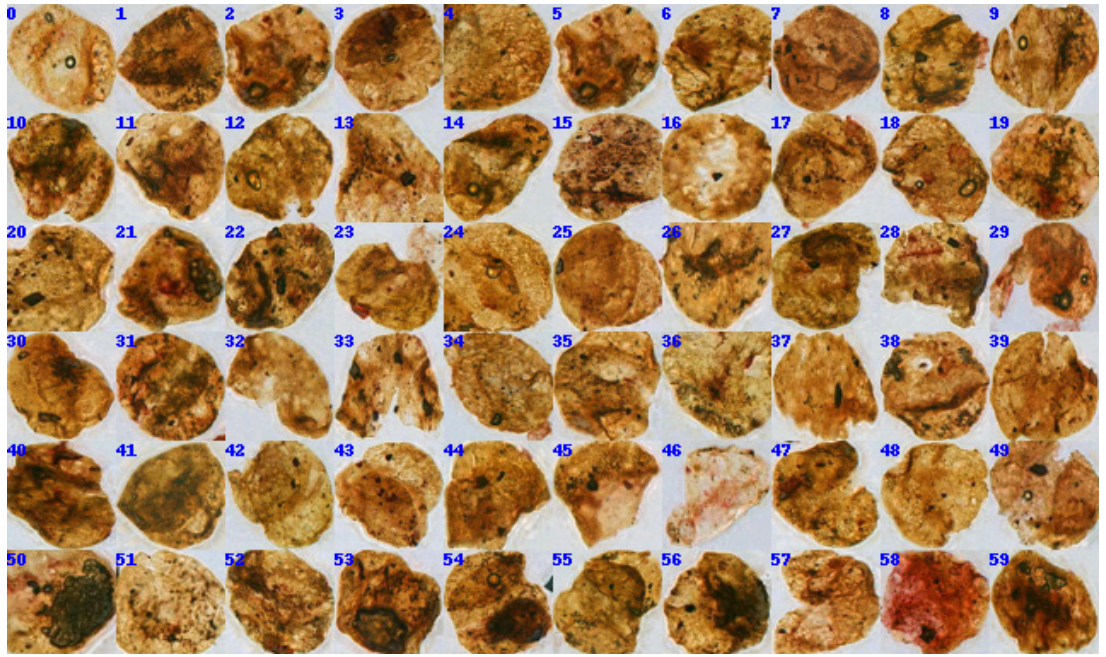
Deltoidospora toralis



Botryococcus

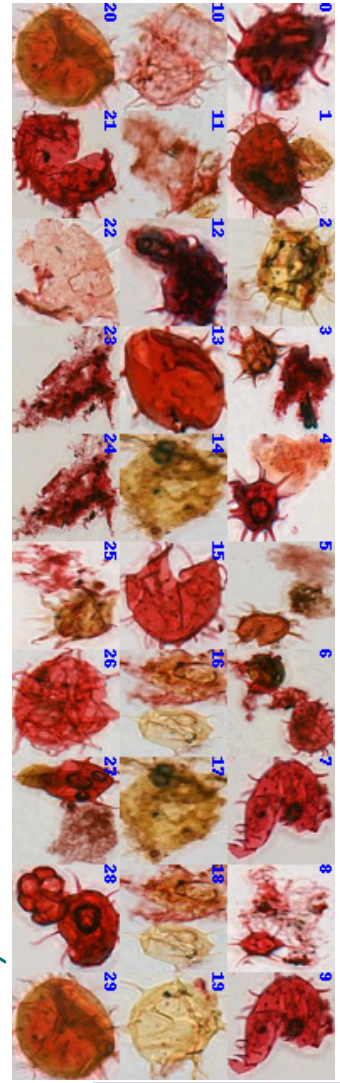


Bisaccate



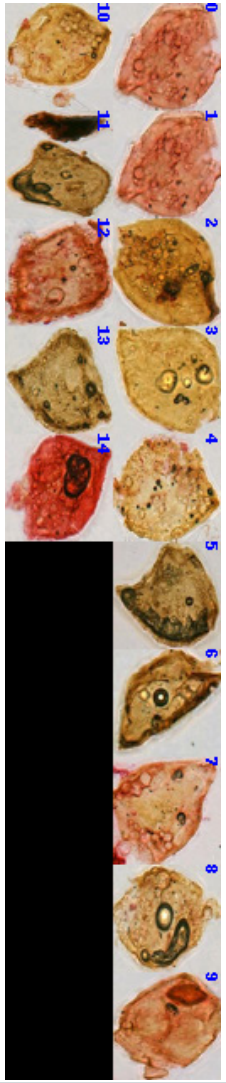
Well B – query results

Micrhystridium

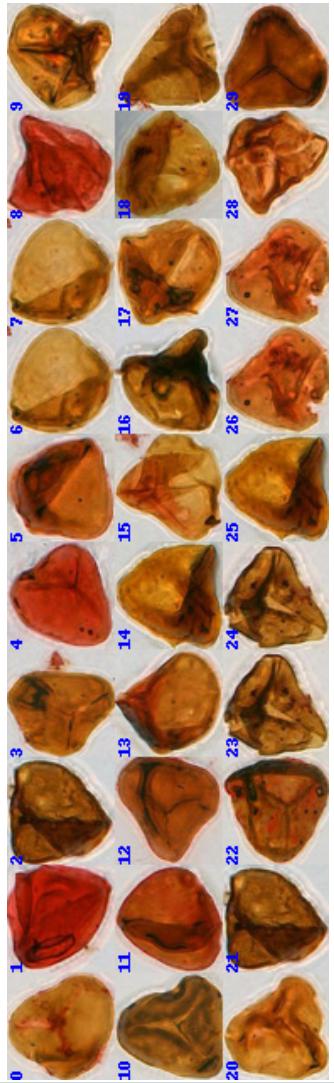


Limbicysta bjaerkei

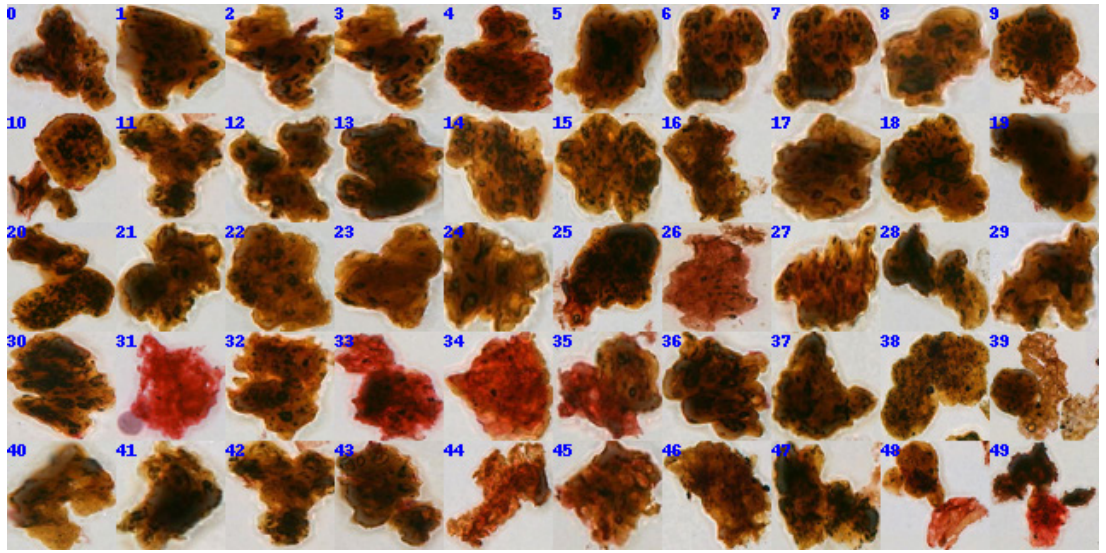
Nannoceratopsis spp.



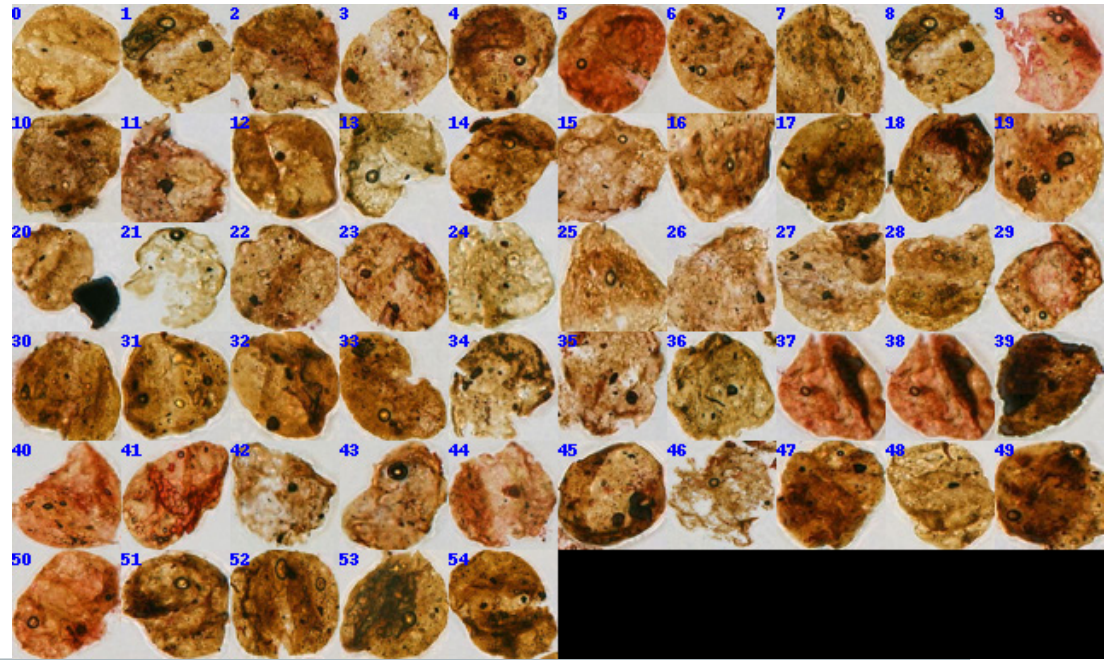
Deltoidospora toralis



Botryococcus



Bisaccate

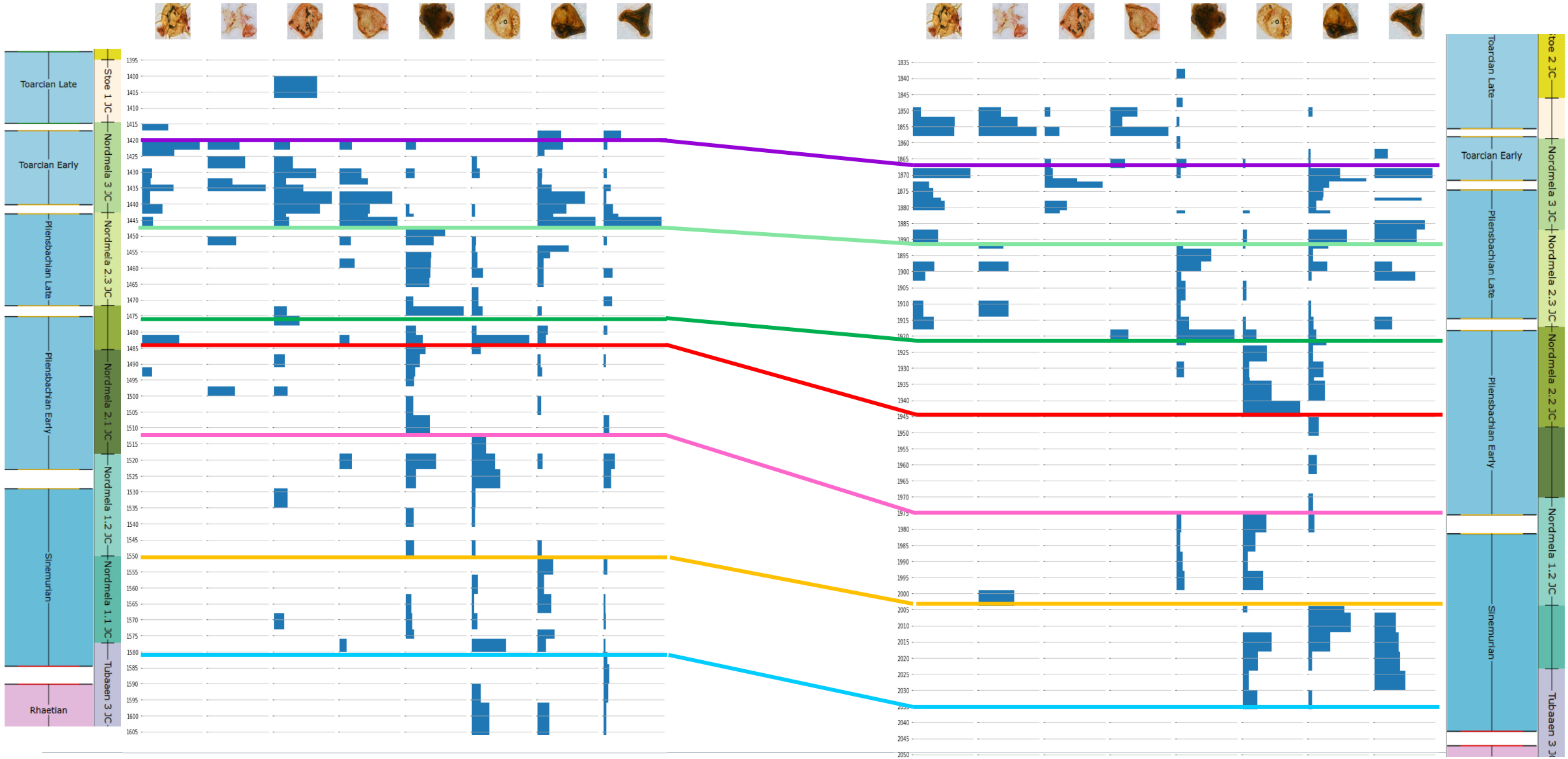


Scampi charts

Well A

Palynologists finds some interesting images to query with - **and** refines queries **and** QCs results,

Well B



Implementation

One webapp for all:

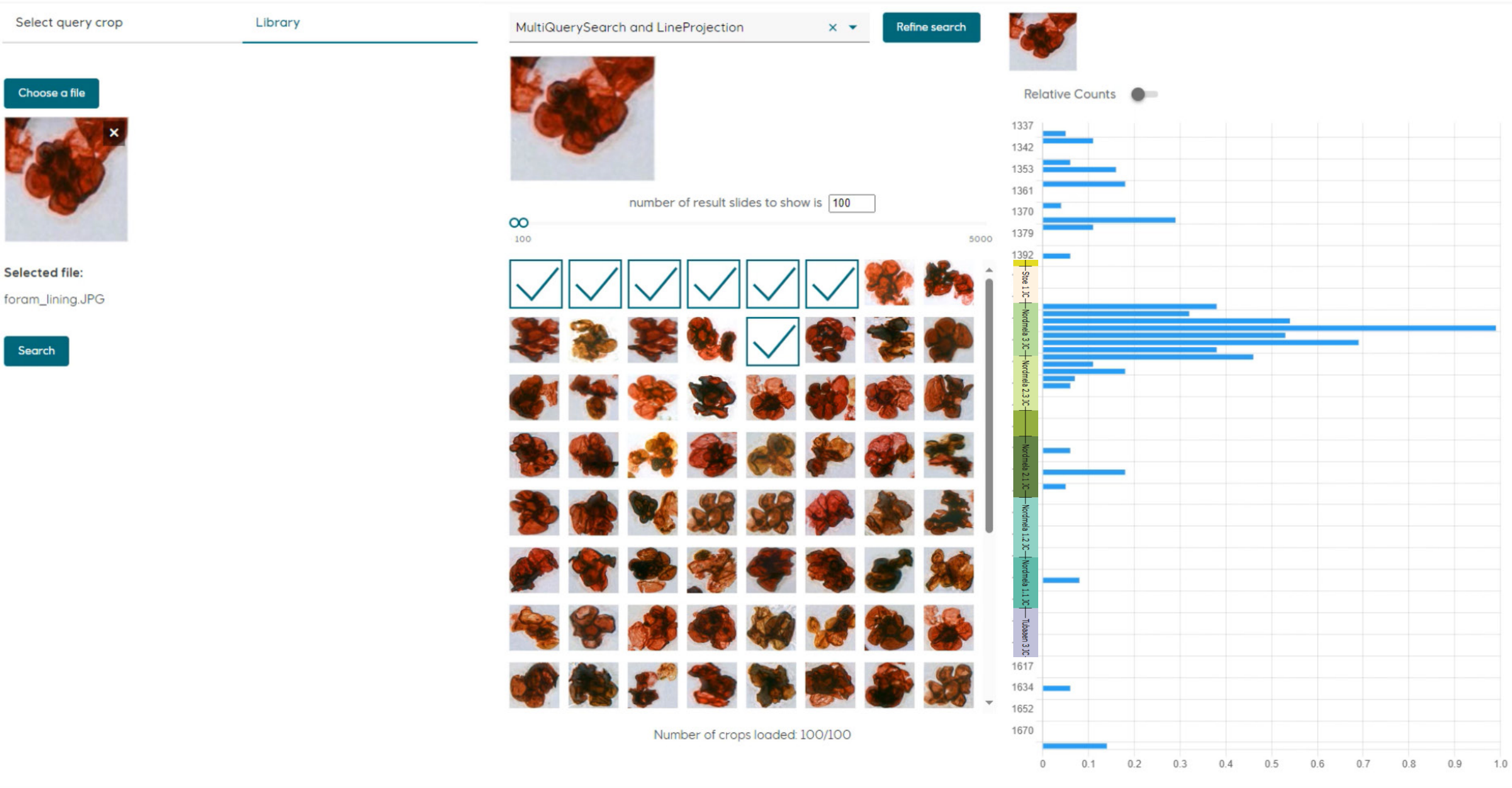
- Logging contractors
- Equinor biostrat team in Shared Services
- Equinor generalists

Select query crop | Library

Choose a file

Drop your file here

Search


Select query crop
Library
Choose a file
Selected file:
foram_lining.JPG
Search
MultiQuerySearch and LineProjection
Refine search
number of result slides to show is
100
5000
Number of crops loaded: 100/100
Relative Counts
1337
1342
1353
1361
1370
1379
1392
1617
1634
1652
1670
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

Select query crop

No well selected

NO 16/2-10

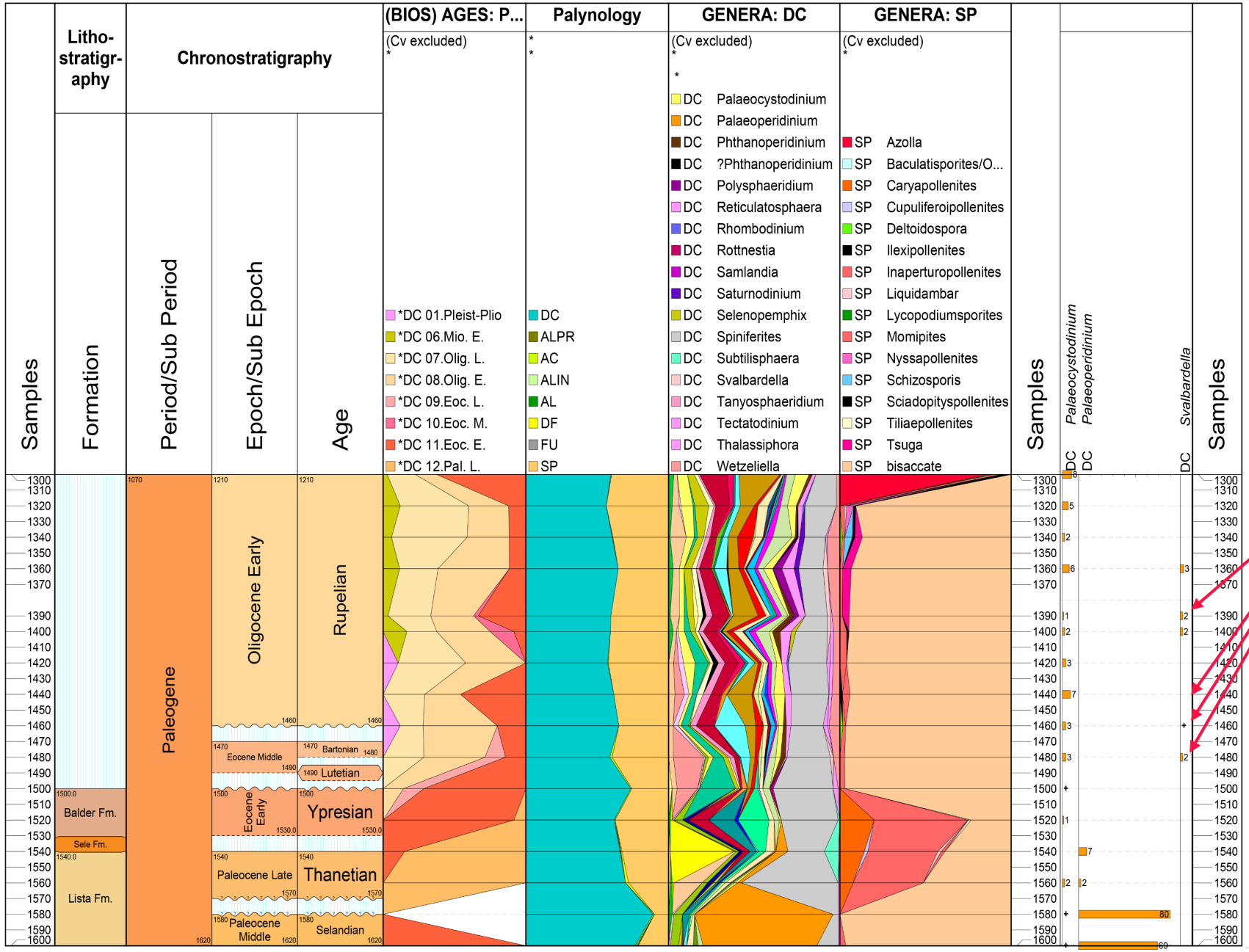
NO 16/2-12

NO 33/9-6

NO 34/10-30

NO 34/2-2

NO 34/2-3


Samples
Litho-stratigr-aphy
Formation
Chronostratigraphy
Period/Sub Period
Epoch/Sub Epoch
Age
Paleogene
Oligocene Early
Rupelian
Eocene Middle
Bartonian
Lutetian
Eocene Early
Ypresian
Paleocene Late
Thanetian
Paleocene Middle
Selandian
Balder Fm.
Sele Fm.
Lista Fm.
(BIOS) AGES: P...
(Cv excluded)
*DC 01.Pleist-Plio
*DC 06.Mio. E.
*DC 07.Olig. L.
*DC 08.Olig. E.
*DC 09.Eoc. L.
*DC 10.Eoc. M.
*DC 11.Eoc. E.
*DC 12.Pal. L.
Palynology
DC
ALPR
AC
ALIN
AL
DF
FU
SP
GENERA: DC
DC Palaeocystodinium
DC Palaeoperidinium
DC Phthanoperidinium
DC ?Phthanoperidinium
DC Polysphaeridium
DC Reticulatosphaera
DC Rhombodinium
DC Rottnestia
DC Samlandia
DC Saturnodinium
DC Selenopemphix
DC Spiniferites
DC Subtilisphaera
DC Svalbardella
DC Tanyosphaeridium
DC Tectatodinium
DC Thalassiphora
DC Wetzeliella
GENERA: SP
SP Azolla
SP Baculatisporites/O...
SP Caryapollenites
SP Cupuliferoipollenites
SP Deltoidospora
SP Ilexipollenites
SP Inaperturopollenites
SP Liquidambar
SP Lycopodiumsporites
SP Momipites
SP Nyssapollenites
SP Schizosporis
SP Sciadopityspollenites
SP Tiliaepollenites
SP Tsuga
SP bisaccate
Palaeocystodinium
Palaeoperidinium
Svalbardella
Absolute Counts

Select query crop | Library

MultiQuerySearch and LineProjection

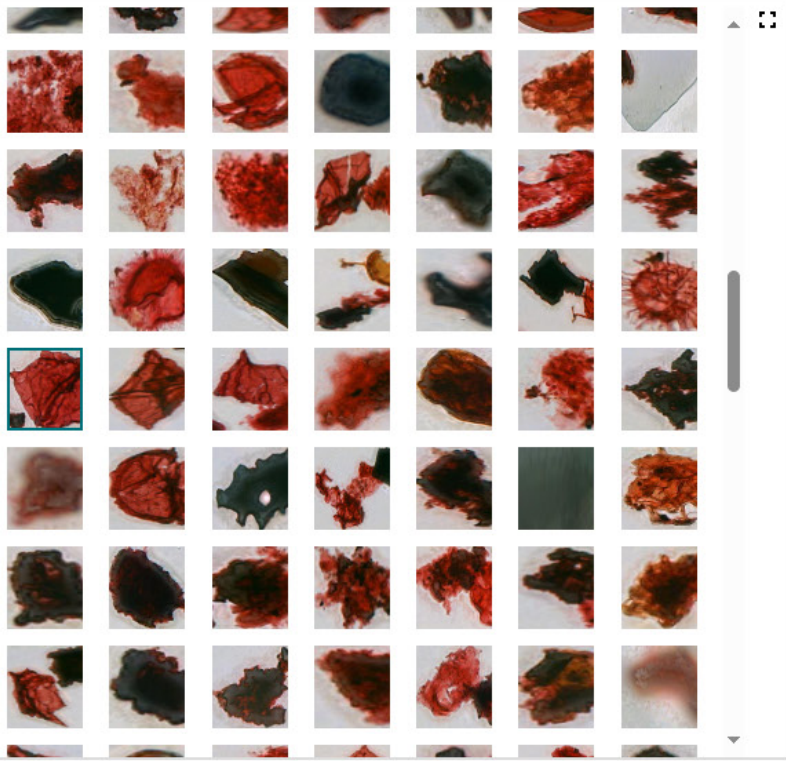
Refine search

1520_mDC

1540_mDC

1560_mDC

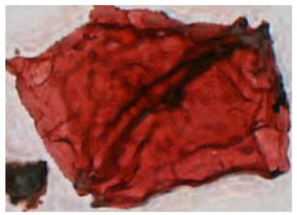
1580_mDC



1600_mDC

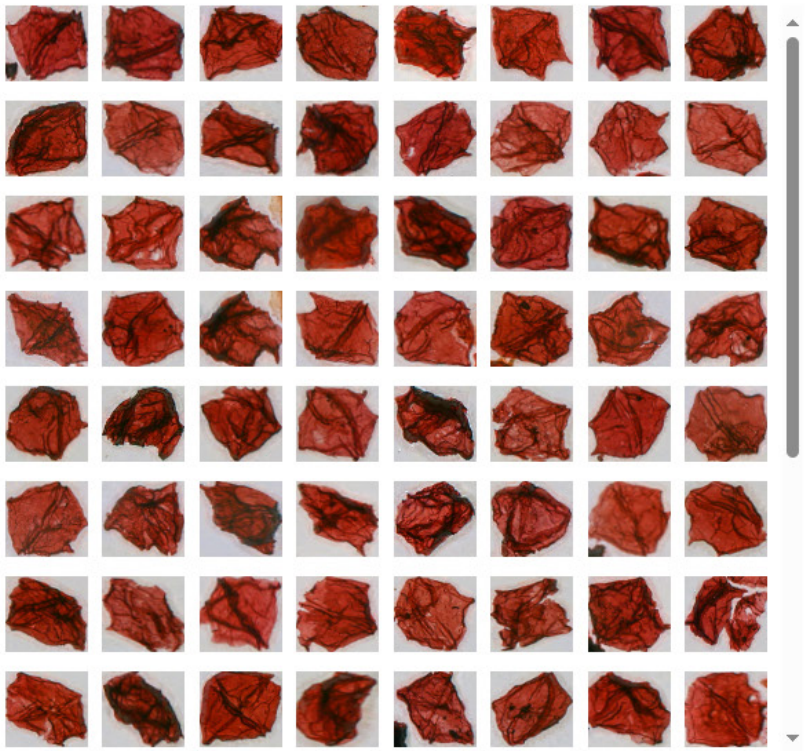
1620_mDC

1633_mDC



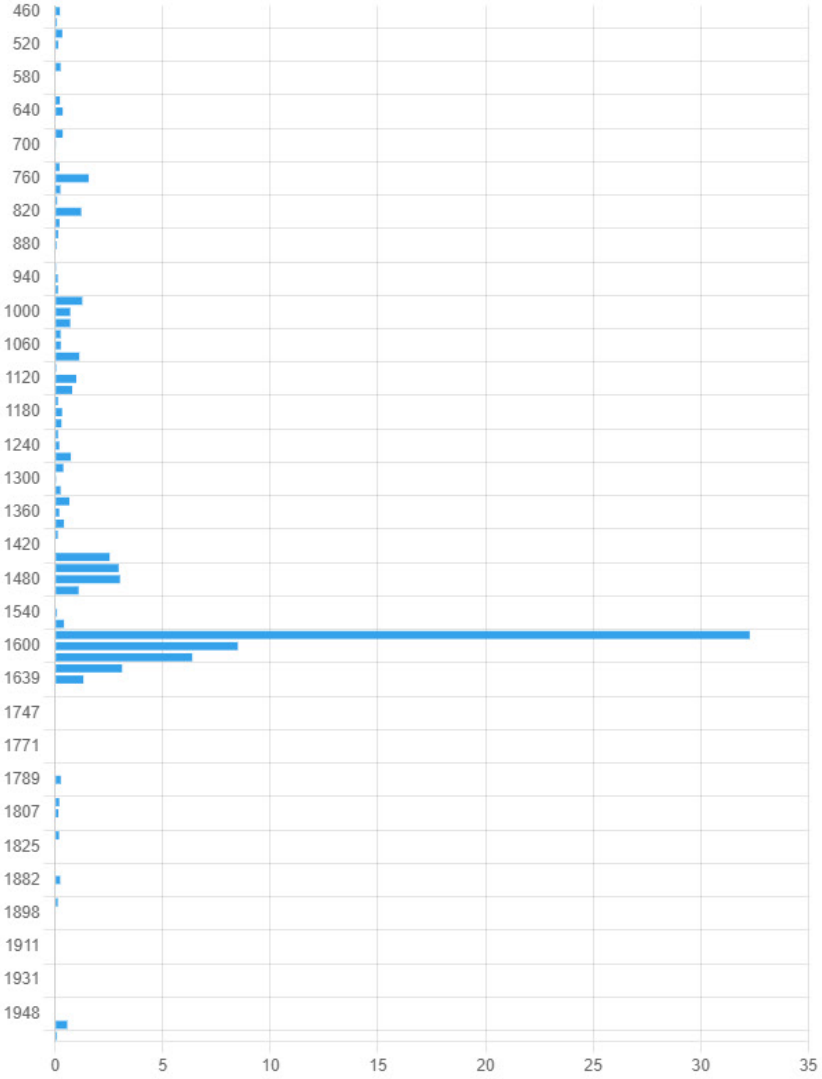
number of result slides to show is 1000

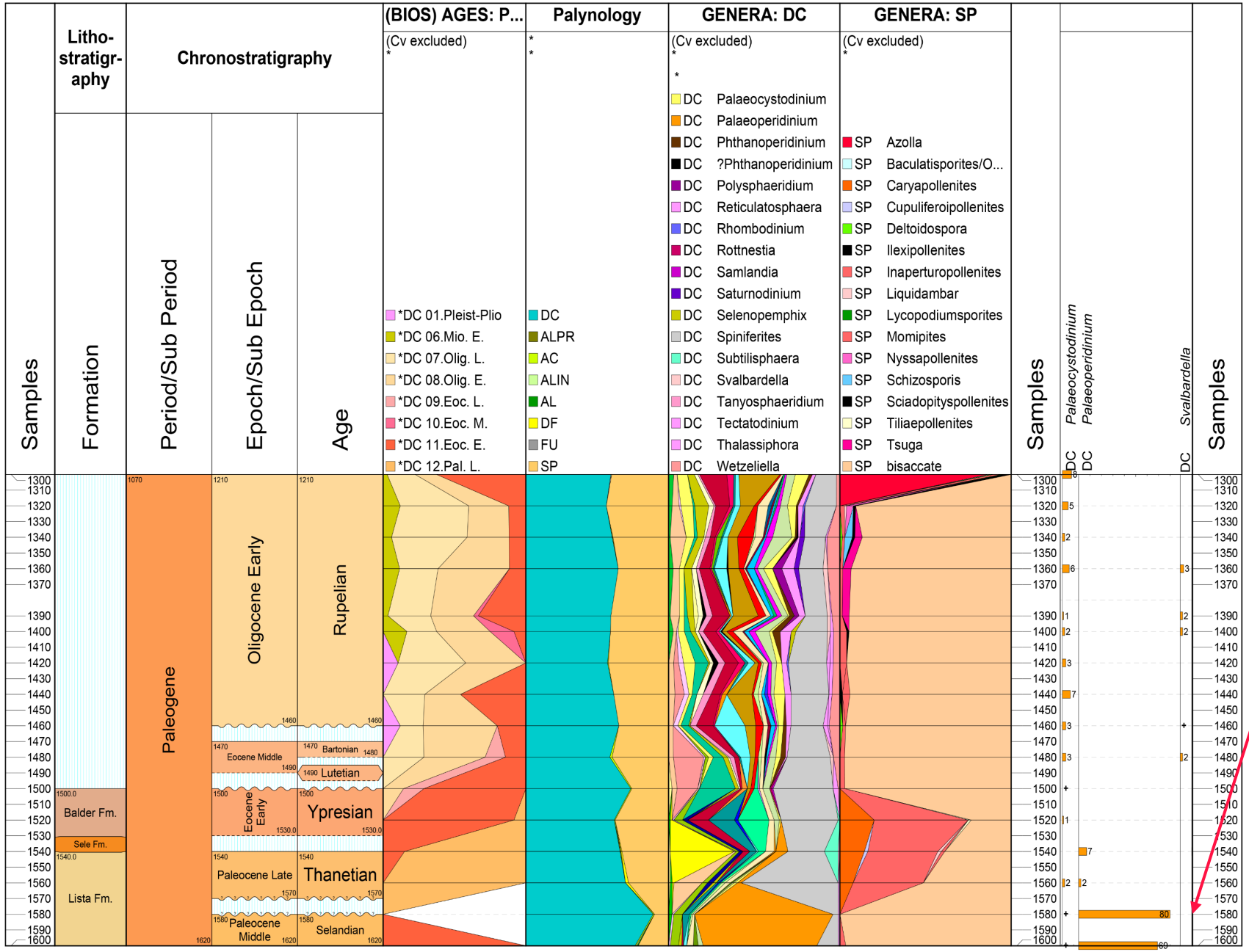
0 1000 5000



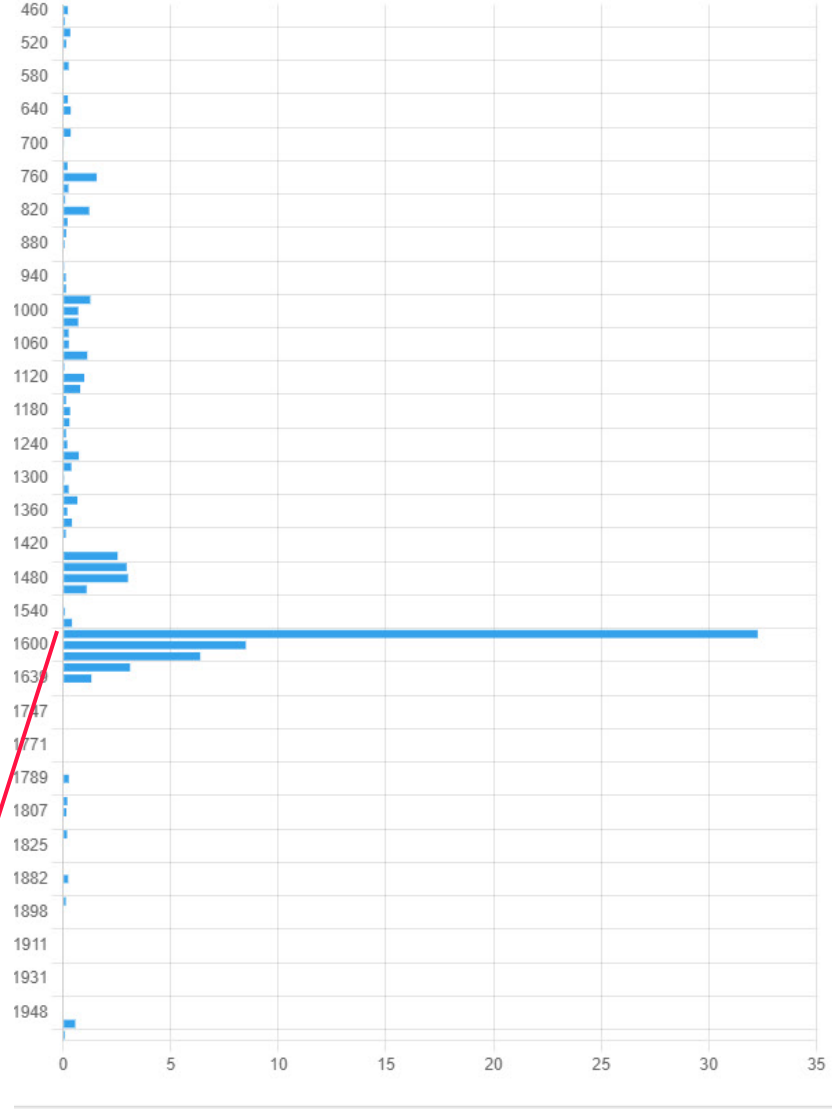
Number of crops loaded: 100/1000

Relative Counts





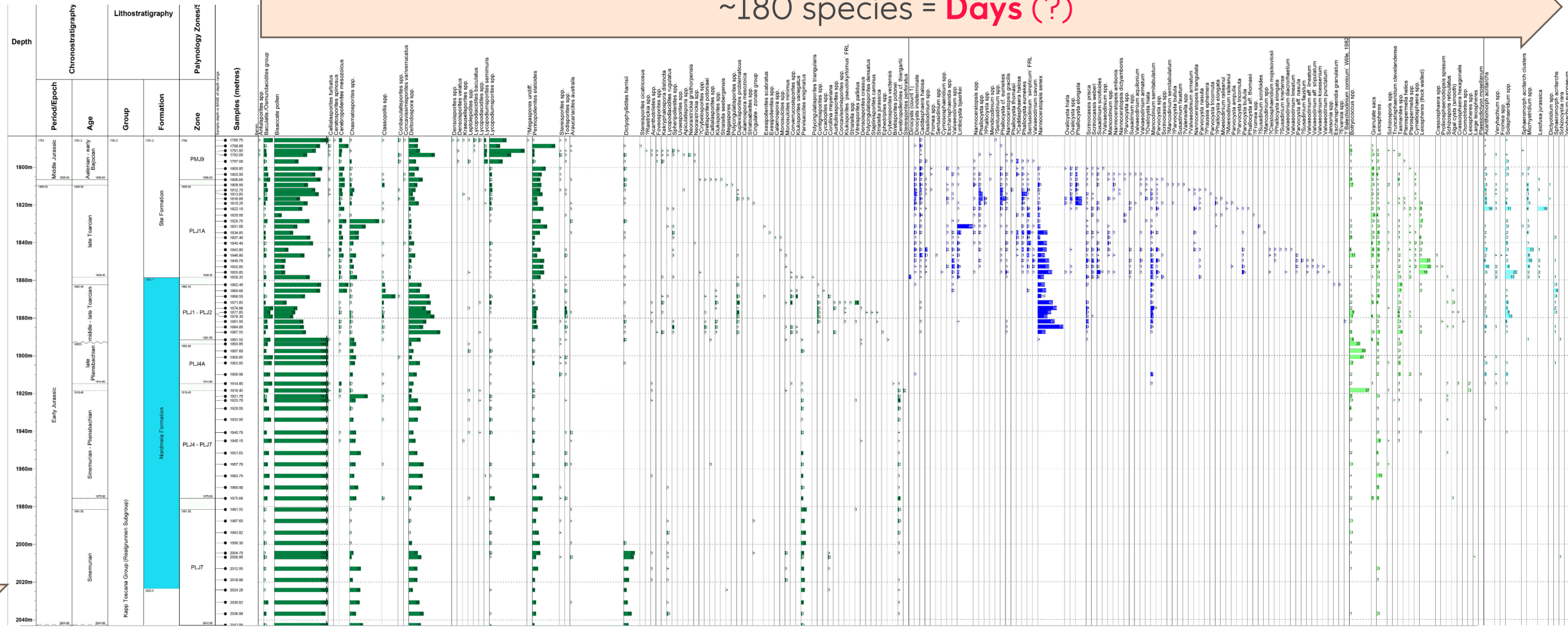
Relative Counts



Imagine creating a distribution chart for a new well…

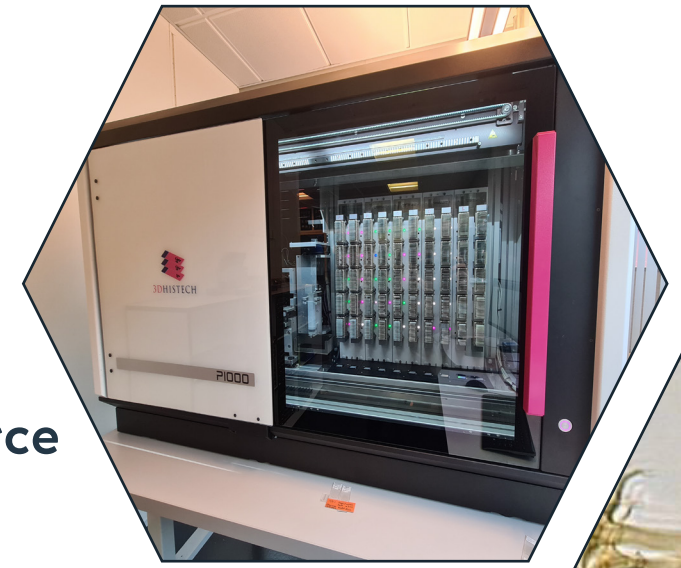
~180 species = **Days** (?)

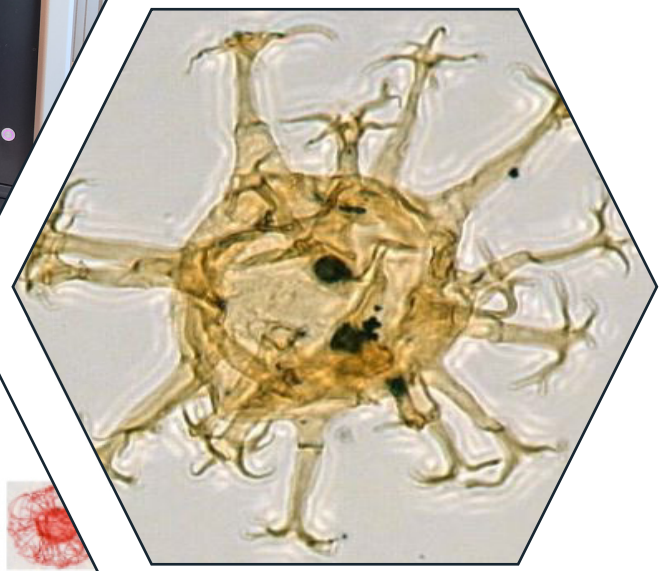
186 slides = **Months**

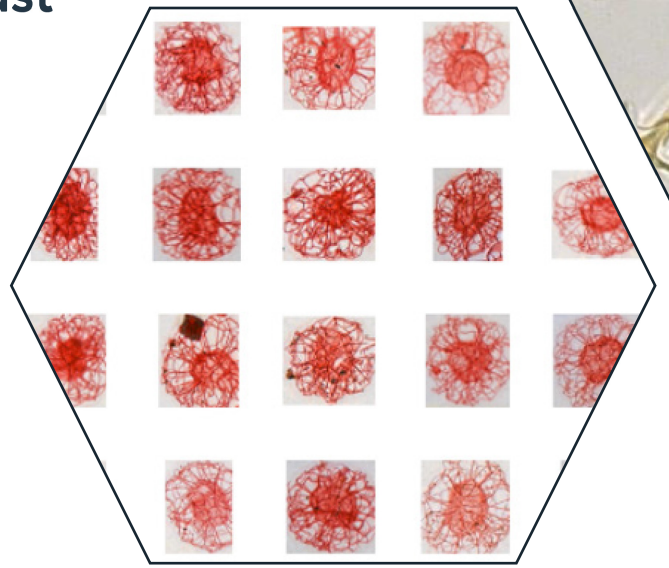


Summary

- NOD's Digital Palynology is an amazing resource
- Scampi enables us to maximize value from it, fast
- Totally new way to view and use biostrat data
- Improves communication with non-specialists







Species Classification Automation for Microfossil Photomicrograph Images (Scampi)

David Wade*, Sissa Stefanowicz, Alex Cullum & Erik Anthonissen
Muhammad Gibran Alfarizi, Håkon Ruud,
Thomas Elvestad‡, Einar Salomonsen‡, Thomas Karlsson‡
‡ Bouvet ASA