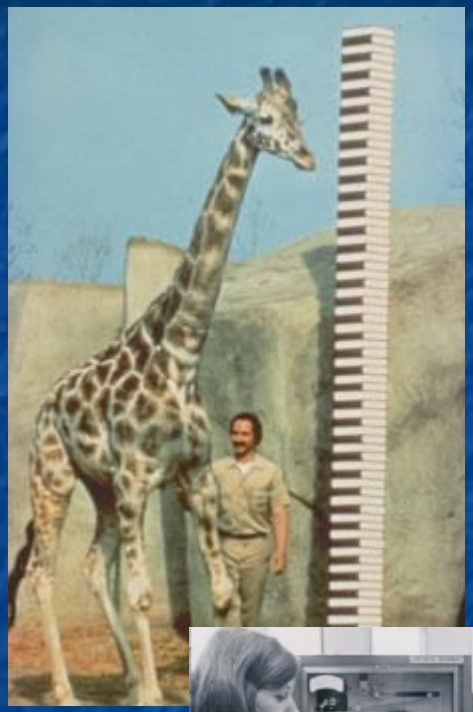
How to Search and FIND – Deep I

Helle Lauridsen
Technology Manager
Proquest CSA
UK
hlauridsen@csa.com

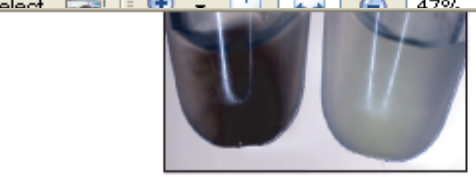
Information explosion



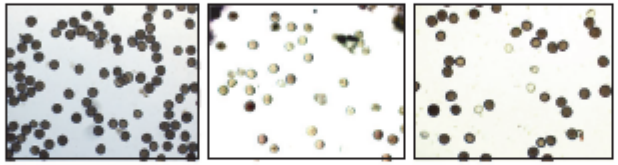
A little History:

- 19th century: Too many journals published to keep track = the first abstracting indexes
- 1907: Chemical abstracts vol.1. contains less than 12,000 abstracts
- 1964 Citation indexing is invented by Eugene Garfield
- 1970'ies first online databases
- 1990'ies WWW becomes a common tool
- 21st century advent of new A&Is – Google Scholar, Live Search Academic, Scopus…
- But – they still search only in the text NOT in the most vital information
- 2007 DEEP INDEXING of Article Images

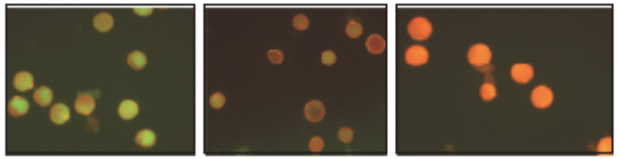
Manual indexing



(a)

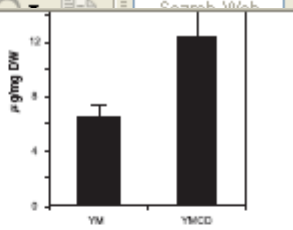


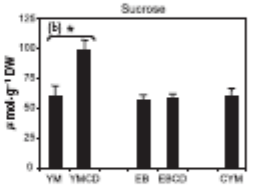
(b)

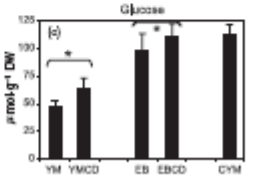


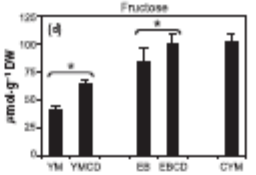
(c)

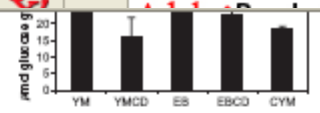
Figure 2. Effects of cold treatment on starch filling and pollen viability. (a) KI/I$_2$ staining of pollen from control plants and plants cold-treated at YM. The difference between control and cold-treated pollen grains is especially clear when pollen grains are pelleted in test tubes. (b) Microscopic observation shows the presence of starch (dark blue colour) in pollen from control plants and the reduction in the number of starch-filled pollen grains after YM cold treatment. At EB, cold treatment caused only a slight reduction (11.9 % ± 1.1 %) in the number of starch-filled pollen. (c) Viability staining, showing the reduction in pollen viability after cold treatment at YM. As a positive control, non-cold-treated YM pollen grains were used ('Control'); boiled pollen grains were used as a negative control ('Boiled'). Green pollen grains are viable; orange pollen grains are not viable.

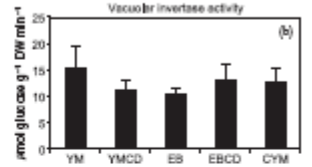












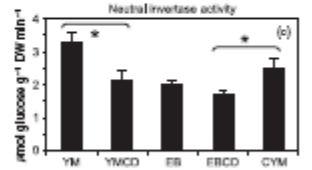


Figure 4. Effect of cold treatment on invertase activity in rice anthers. Invertase activity was assayed in YM (YM and YMCD), EB (EB and EBCD), and CYM anthers. Error bars represent standard errors and *t*-tests were used to test the significance of the differences caused by the cold treatment (YM/YMCD, EB/EBCD, EB/CYM). The Student method with an asterisk (*) indicates significant differences ($P < 0.05$).

Figure 3. Effect of cold treatment on sugar levels in rice anthers. Levels of total non-reducing sugars (a), sucrose (b), glucose (c) and fructose (d) in anthers harvested at YM (no-treated control, YM; cold-treated, YMCD) and EB (no-treated control, EB; cold-treated, EBCD), and anthers cold-treated at YM and then allowed to develop under normal conditions until EB (CYM). The non-reducing sugar results are the average of two biological repeat experiments. Error bars represent standard errors and the *t*-test was used to test the effect of the cold treatment (YM/YMCD and EB/EBCD comparisons); brackets labelled with an asterisk (*) indicate significant differences ($P < 0.05$). FW, fresh weight; DW, dry weight.

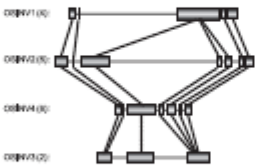
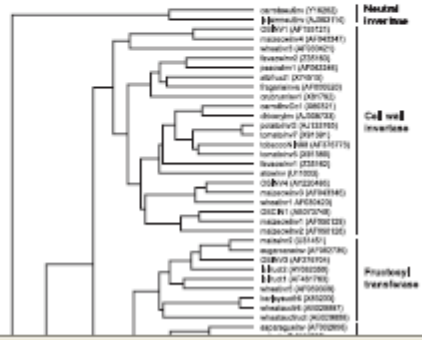


Figure 5. Cloning and structural analysis of four rice invertase genes. (a) Comparison of the intron–exon structure of four rice invertase genes, showing conservation of the intron positions (dotted lines) between *OSINV2* and 4 (cell wall invertase), *OSINV1* (vacuolar invertase), and *OSINV3* (fructosyl transferase). The number of introns in each gene is shown in brackets. (b) Amino acid



Why Index Tables And Figures?

- They contain important and valuable information
- Figures and tables represent the distilled essence of research – the closest thing to raw datasets
- Researchers want access to data
- They are invisible

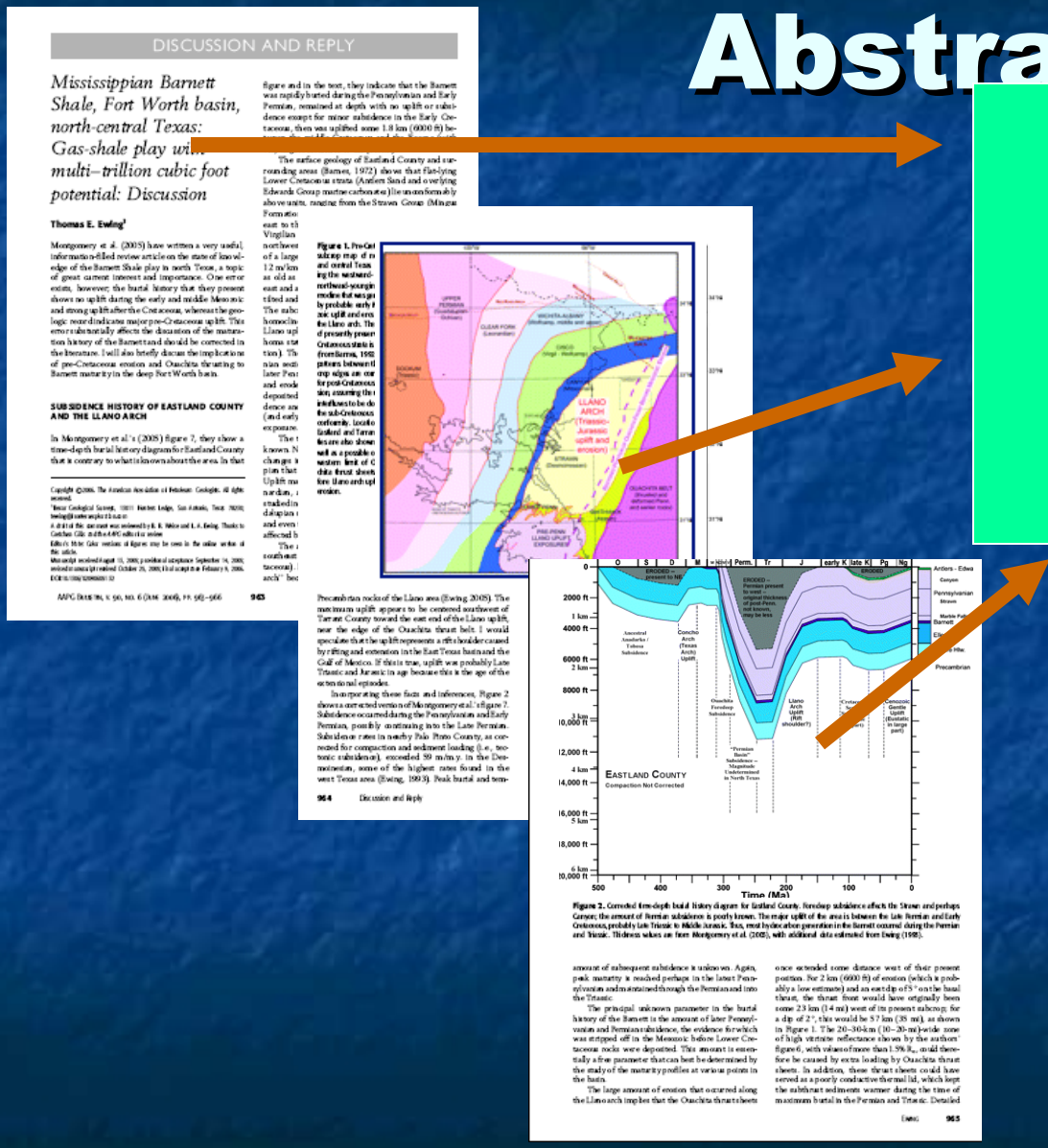
Reasons Why Data Are Hidden In Traditional Searches

- Data **variables** do not appear in any index.
  - there are no indexing 'hooks' in title, abstract or caption for "*dissolved oxygen*", below.
2. A search of the full text bypasses the image files
   - text in tables & figures is considered an image, not searchable text

Sts.	Depth (m)	Sal.	Temp. (°C)	pH	DO ($mg L^{-1}$)	Gravel (%)	Sand (%)	Silt (%)	Clay
1	2.1	31.8	24.4	8.66	6.92	0	99.9	0.1	0
2	2.5	31.9	24.3	8.67	7.05	0	95.8	3.2	1
3	1.8	31.6	24.5	8.63	7.00	0	99.8	0.2	
4	1.7	31.7	25.6	8.68	7.06	0	90.9	7.8	
5	2.0	31.7	25.5	8.66	6.76	0	8.4	66.3	
6	2.7	32.2	25.5	8.70	6.90	0	7.5		
7	1.9	31.9	25.2	8.67	7.02	0			
8	3.0	31.9	24.3	8.61	6.71				
9	4.3	32.2	24.1	8.65	6.37				
10	2.3	31.8	24.6						

Table 1. Depth, physico-chemical and sedimentological variables.

Deep Indexing – Abstract Record



DISCUSSION AND REPLY

Mississippian Barnett Shale, Fort Worth basin, north-central Texas: Gas-shale play with multi–trillion cubic foot potential: Discussion

Thomas E. Ewing[1]

Montgomery et al. (2005) have written a very useful, information-filled review article on the state of knowledge of the Barnett Shale play in north Texas, a topic of great current interest and importance. One error exists, however; the burial history that they present shows no uplift during the early and middle Mesozoic and strong uplift after the Cretaceous, whereas the geologic record indicates major pre-Cretaceous uplift. This error substantially affects the discussion of the maturation history of the Barnett and should be corrected in the literature. I will also briefly discuss the implications of pre-Cretaceous erosion and Ouachita thrusting to Barnett maturity in the deep Fort Worth basin.

SUBSIDENCE HISTORY OF EASTLAND COUNTY AND THE LLANO ARCH

In Montgomery et al.'s (2005) Figure 7, they show a time-depth burial history diagram for Eastland County that is contrary to what is known about the area. In that figure and in the text, they indicate that the Barnett was rapidly buried during the Pennsylvanian and Early Permian, remained at depth with no uplift or subsidence except for minor subsidence in the Early Cretaceous, then was uplifted some 1.8 km (6000 ft) be-

The surface geology of Eastland County and surrounding areas (Barnes, 1972) shows that flat-lying Lower Cretaceous strata (Antlers Sand and overlying Edwards Group marine carbonates) lie unconformably above units, ranging from the Strawn Group (Missis-

Precambrian rocks of the Llano area (Ewing, 2005). The maximum uplift appears to be centered southwest of Tarrant County toward the east end of the Llano uplift, near the edge of the Ouachita thrust belt. I would speculate that the uplift represents a rift shoulder caused by rifting and extension in the East Texas basin and the Gulf of Mexico. If this is true, uplift was probably Late Triassic and Jurassic in age because this is the age of the extensional episodes.

In incorporating these facts and inferences, Figure 2 shows a corrected version of Montgomery et al.'s Figure 7. Subsidence occurred during the Pennsylvanian and Early Permian, possibly continuing into the Late Permian. Subsidence rates in nearby Palo Pinto County, as corrected for compaction and sediment loading (i.e., tectonic subsidence), exceeded 50 m/m.y. In the Delaware basin, some of the highest rates found in the west Texas area (Ewing, 1993). Peak burial and tem-

964 Discussion and Reply

EASTLAND COUNTY
Compaction Not Corrected

Time (Ma)

Figure 2. Corrected time-depth burial history diagram for Eastland County. Foreloop subsidence affects the Strawn and perhaps Canyon; the amount of Permian subsidence is poorly known. The major uplift of the area is between the Late Permian and Early Cretaceous, probably Late Triassic to Middle Jurassic. Thus, most hydrocarbon generation in the Barnett occurred during the Permian and Triassic. Thickness values are from Montgomery et al. (2005), with additional data estimated from Ewing (1993).

amount of subsequent subsidence is unknown. Again, peak maturity is reached perhaps in the latest Pennsylvanian and remained through the Permian and into the Triassic.

The principal unknown parameter in the burial history of the Barnett is the amount of later Pennsylvanian and Permian subsidence, the evidence for which was stripped off in the Mesozoic before Lower Cretaceous rocks were deposited. This amount is essentially a free parameter that can best be determined by the study of the maturity profiles at various points in the basin.

The large amount of erosion that occurred along the Llano arch implies that the Ouachita thrust sheets once extended some distance west of their present position. For 2 km (6600 ft) of erosion (which is probably a low estimate) and an eastdip of 5° on the basal thrust, the thrust front would have originally been some 23 km (14 mi) west of its present subcrop, for a dip of 2°; this would be 57 km (35 mi), as shown in Figure 1. The 20–30-km (10–20-mi)-wide zone of high vitrinite reflectance shown by the authors' figure 6, with values of more than 1.5% R_o, could therefore be caused by extra loading by Ouachita thrust sheets. In addition, these thrust sheets could have served as a poorly conductive thermal lid, which kept the subthrust sediments warmer during the time of maximum burial in the Permian and Triassic. Detailed

Ewing 965

Abstract Record
enhanced with:

- Objects thumbnails
- Captions
- Index terms
- Link to Object DB
- Other metadata

What Researchers Currently Do

- Search for photographs and maps more than tables, figures or graphs
- Use Google Images most often
- Level of satisfaction with traditional searches consistently rated low
- locating objects is “difficult”
- “in general, academic figures, tables, and graphs are not available to search”

From idea to reality

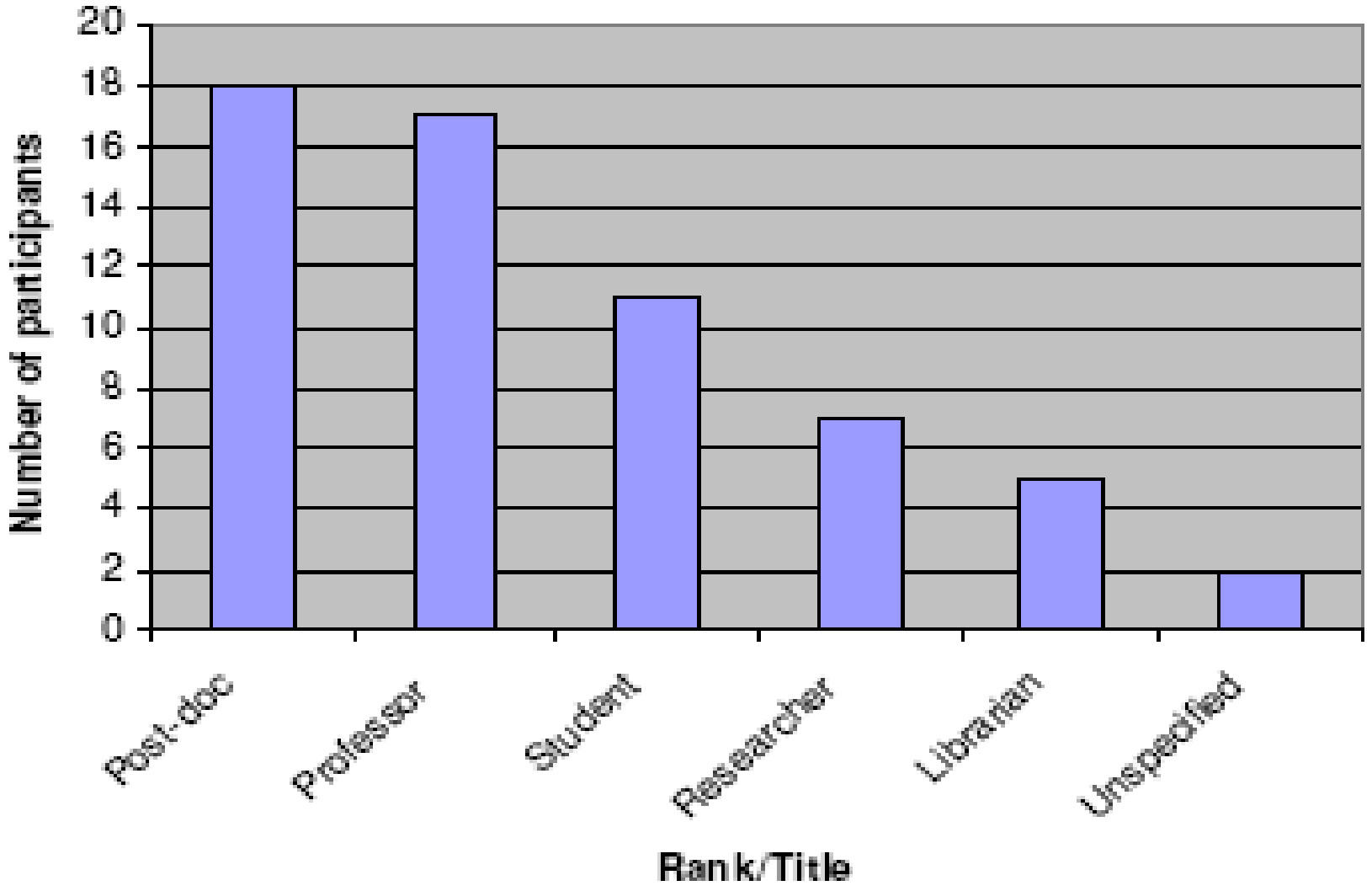
- An innovative Company
- A Prototype database of 325,000 objects
- In depth market research set up by Carol Tenopir from Tennessee University
- 60+ scientists, students and librarians
- Lots of travelling and face to face meetings with scientists
- A White Paper
- Agreements with major publishers

In Depth Market Research: Participants

	Universities	Research Institutes	Totals
United States	5	1	6
Europe	2	1	3
Totals	7	2	9

- ~~9 institutions~~
- 60 scientists (mostly life sciences)
- Over 380 searches

In Depth Market Research: Participants

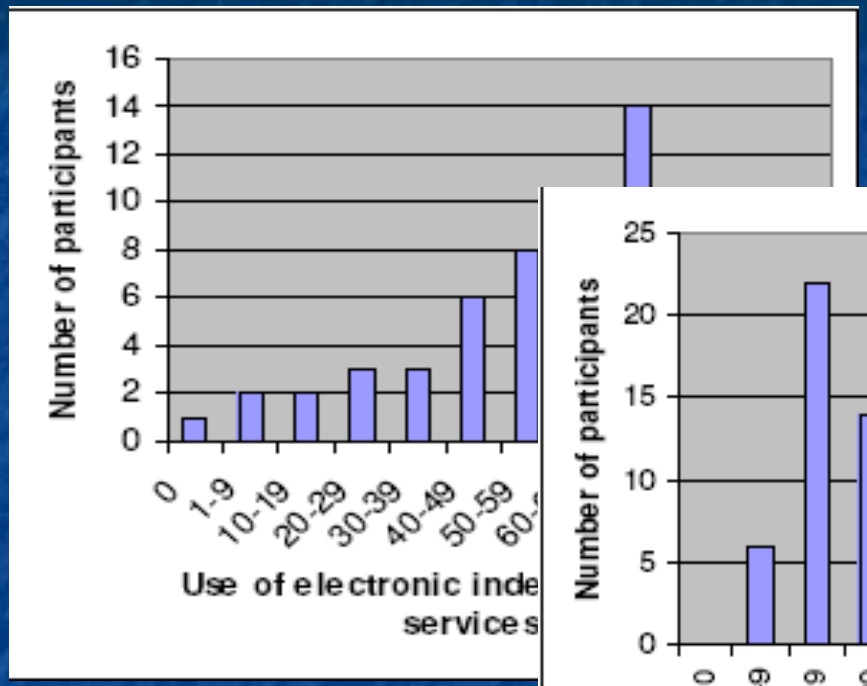


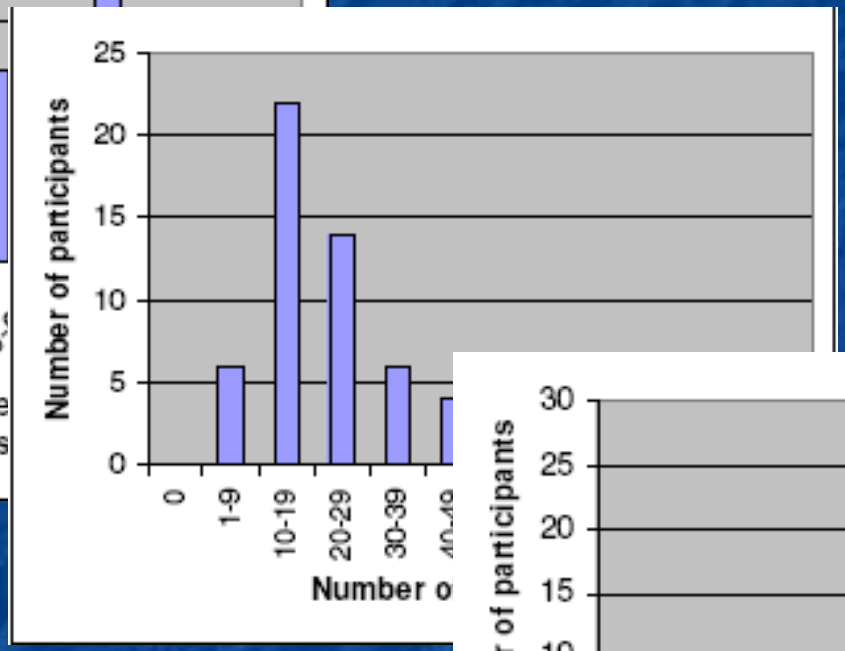
In depth market research

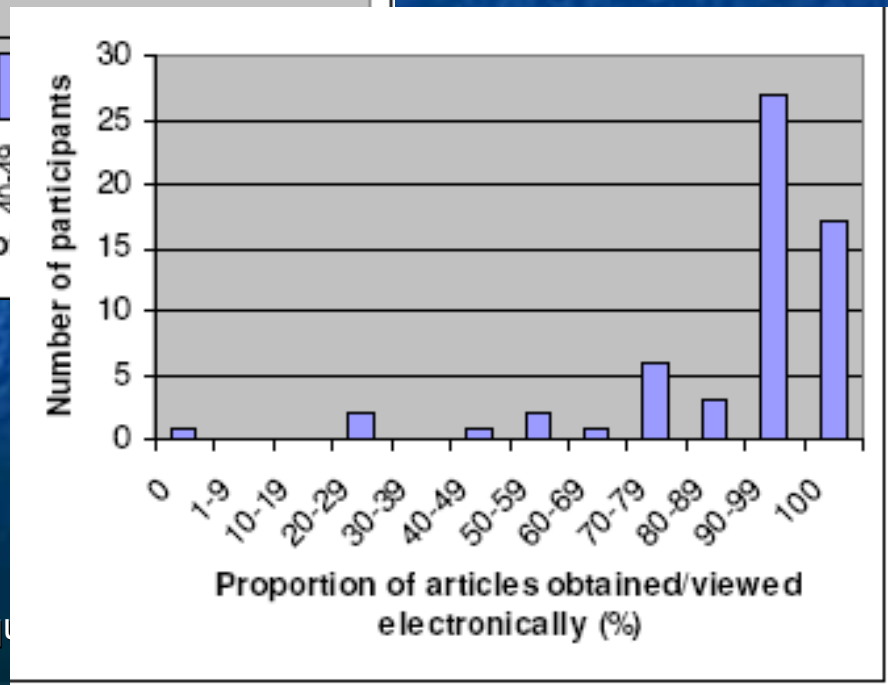
The research team wanted to unveil:

- Current Practices and Experiences
- Expectations for the Tables and Figures Index
- Experiences with Tables and Figures Index
- Effectiveness of Tables and Figures Index

Current Practices and Experiences



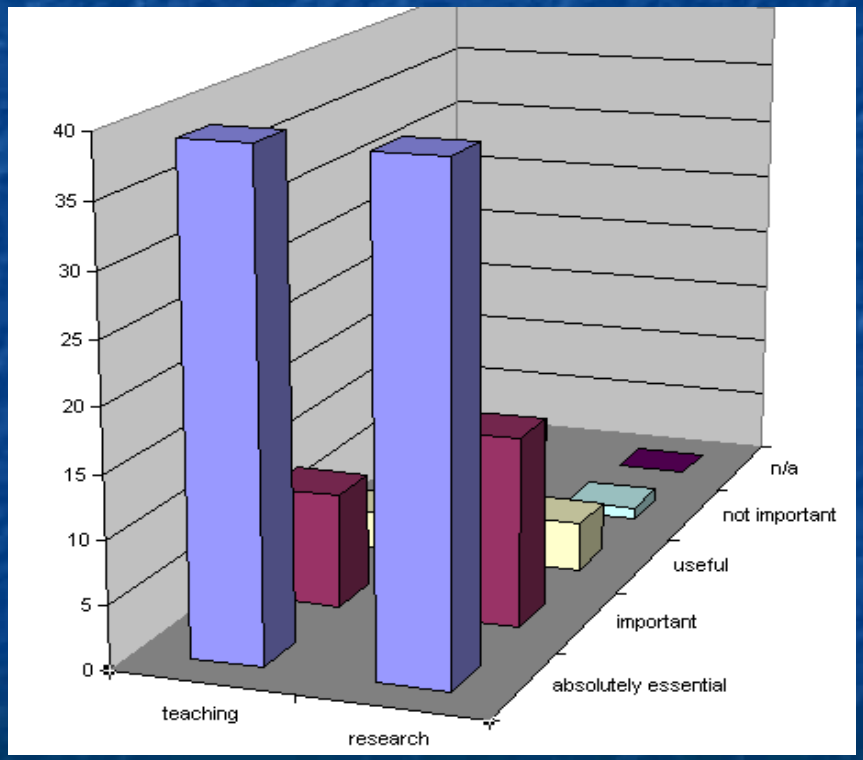




A highly experienced and computer literate test group

Helle Lauridsen, Proqu

Expectations for the Tables and Figures Index



Most of the participants expected the ability of searching in figures as absolutely essential

- **Teaching, lectures, talks, presentations including incorporating tables and figures found directly into presentation software, such as PowerPoint**
- **Locating and retrieving data of particular types, such as tables, graphs, figures, maps and photographs**
- **Making comparisons between one's own work and the work of others as well as comparing the work of multiple other researchers for a variety of purposes; putting one's work into the context of research in the discipline**
- **Gaining faster and more precise understanding of the work reported in other papers by direct examination of the objects embedded in other articles**
- **In support of writing and other forms of scholarly production including conducting meta-analyses and writing review papers, writing journal articles, writing research proposals, developing formulae and models, and generating hypotheses**
- **Faster and more efficient searching, with smaller, more precise results sets**

Experiences with Tables and Figures Index

- . "I can find the tables and figures that I need quickly, [and] it can save me a lot of time. I can work more efficiently" (Post Doc, Biology)
- "It makes the search much quicker when it is focused" (Post Doc, Biology)
- that "the tables and figures are really helpful for scanning large sets of data first" (Post Doc, Oceanography).
- "[i]t takes less time to find the information I want and especially I would find this useful when making a presentation" (Student, Biology).
- "I could find relevant information more quickly and images that were useful for presentations and research" (Professor, Engineering).

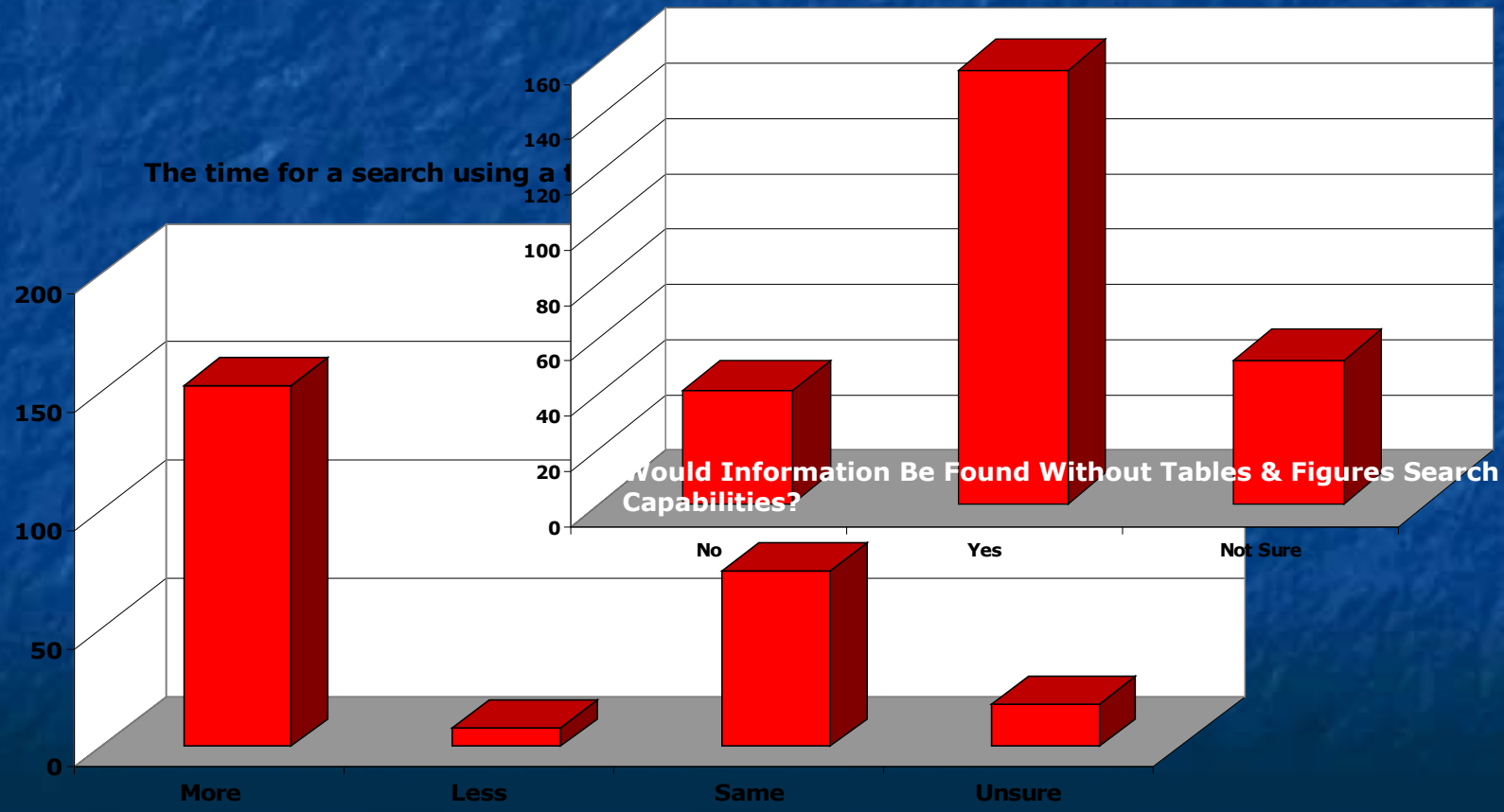
Experiences with Tables and Figures Index

They also told us…

- **Quality of the tables was PARAMOUNT.**
- Rights – with proper attribution tables and figures can be extracted directly from the database and used in teaching and other work.
- Linking to the full text was crucial since they would not use an image unless they were sure of the context.
- They wanted to see a list of articles as well as a list of relevant objects
- Overview at a glance right after searching, no unnecessary clicks

Effectiveness of Tables and Figures Index

Surprisingly, even the small dataset in the prototype revealed the usefulness of a tables and figures index:

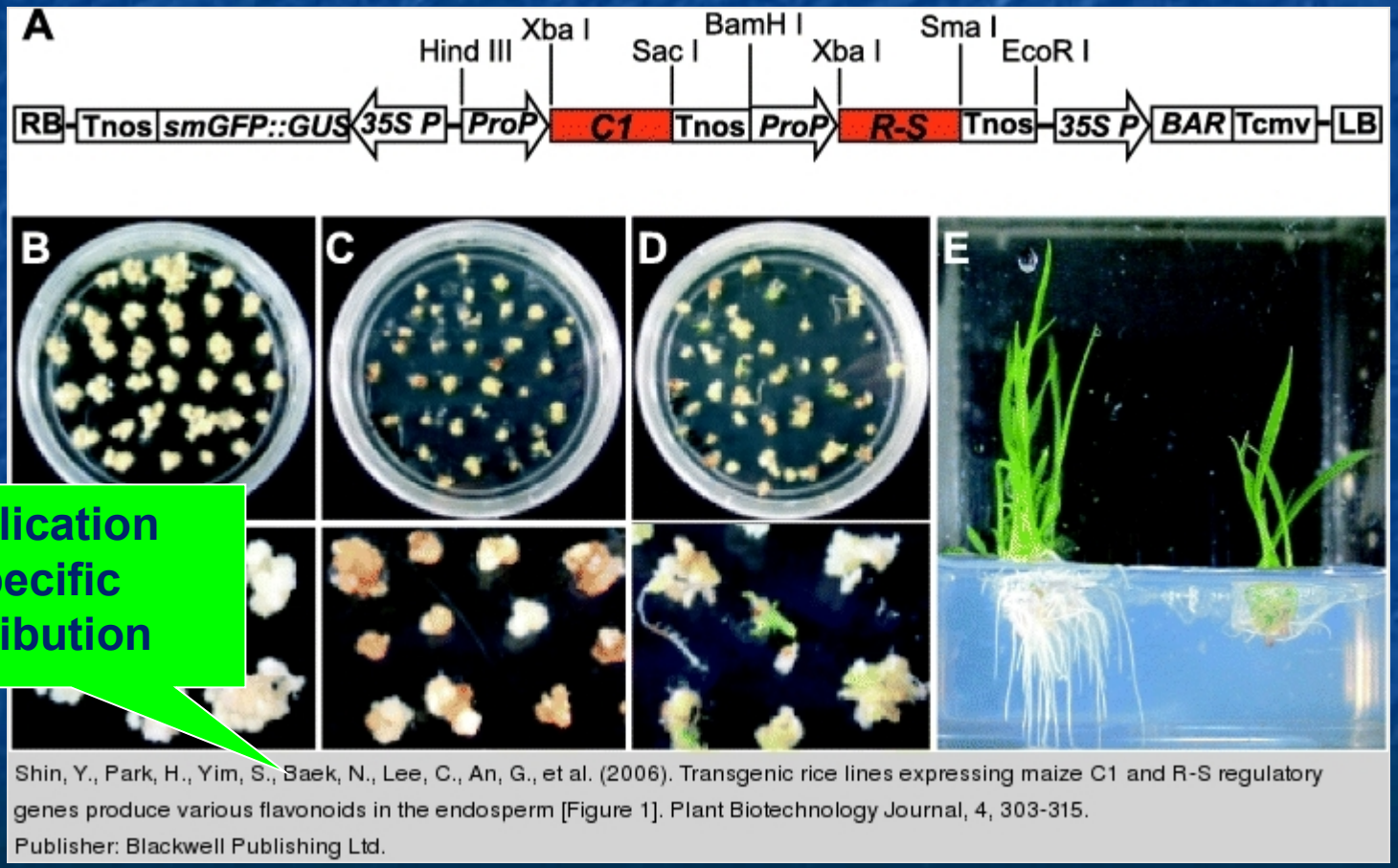


From prototype to reality

The feedback from the market research sent the development team back to the drawing board to make the required changes:

The Product Design Changed

The figure quality improved drastically

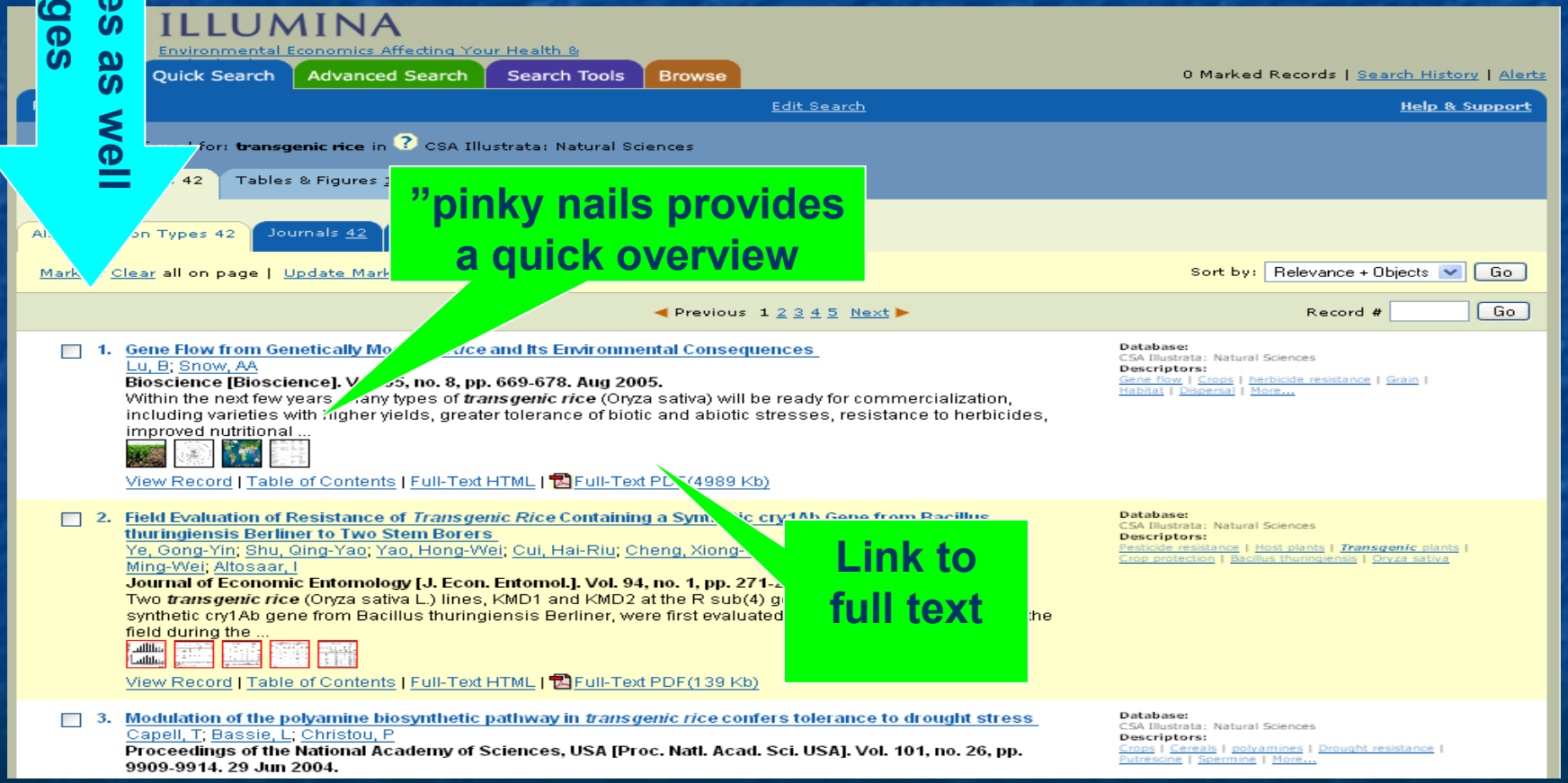


Shin, Y., Park, H., Yim, S., Baek, N., Lee, C., An, G., et al. (2006). Transgenic rice lines expressing maize C1 and R-S regulatory genes produce various flavonoids in the endosperm [Figure 1]. Plant Biotechnology Journal, 4, 303-315. Publisher: Blackwell Publishing Ltd.

Publication specific Attribution

The Product Design Changed – and improved

List of articles as well as images



Clear sharp images + mouseover information = quick overview

Affiliation fax +82-54-279-0659, ymwoo@postech.ac.kr

Source Plant Biotechnology Journal [Plant Biotechnol. J.]. Vol. 4, no. 3,

Notes Figures, 10; tables, 2; references, 30.

Objects

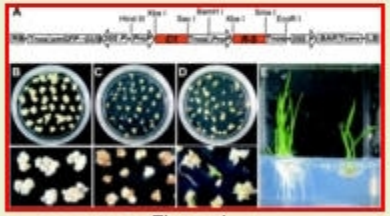


Figure 1.

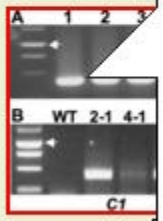


Figure 2.

Caption: Molecular genetic analyses of independent C1 / R-S lines. Arrows indicate 1-kb band of DNA size ladder. (A) Using BAR gene primers, polymerase chain reaction (PCR) products were amplified with genomic DNAs isolated from ***transgenic*** lines 1, 2, 3, 4, 6, 9 and 19. Plasmid DNA (PL) containing transgenes was included as a PCR-positive control, whereas wild-type (WT) genomic DNAs served as a negative control. (B) Reverse transcriptase-polymerase chain reaction (RT-PCR) products show relative expression levels of C1 and R-S transgenes in developing kernels of WT and 2-1, 4-1 and 9-2 T 2 ***transgenic*** lines.

Category: Figure; Photograph; Gel

Object Subject Terms: Genomic DNAs; Plasmid DNA; Reverse transcriptase-polymerase chain reaction; T 2 ***transgenic*** lines

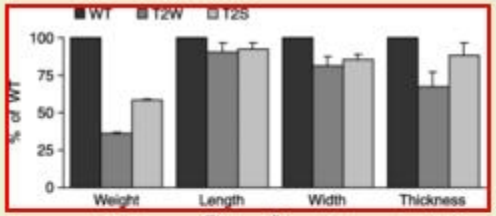


Figure 3.


Fig

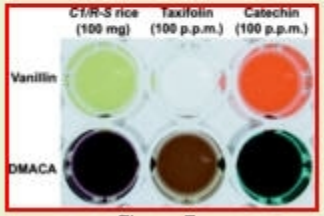


Figure 7.

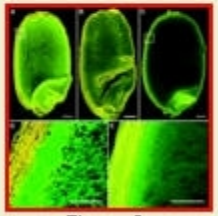


Figure 8.

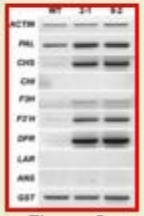


Figure 9.

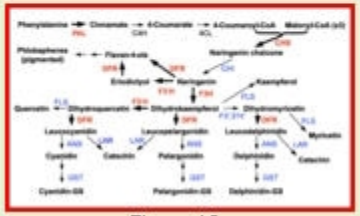
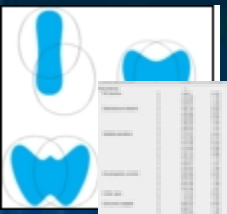
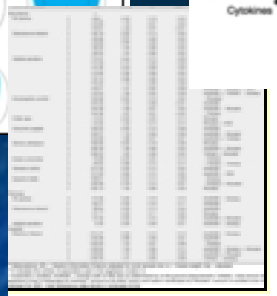
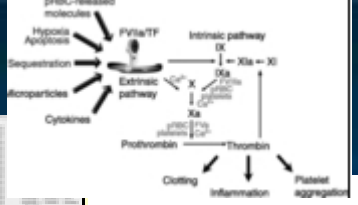
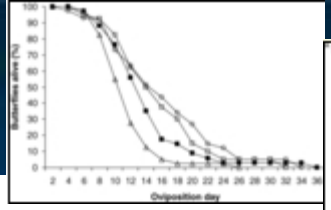
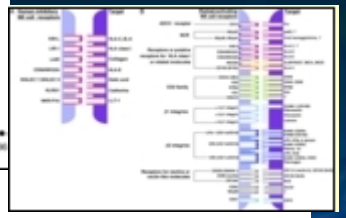
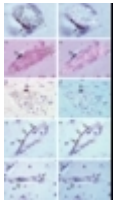


Figure 10.

- as well as original images

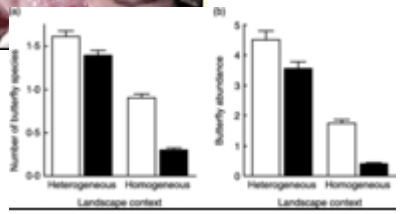




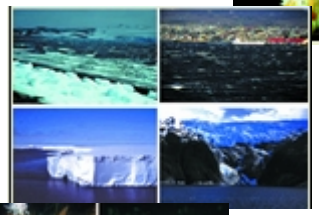
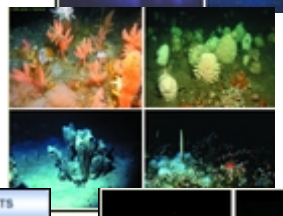


- or just at

http://info.csa.com/csaillustrata/

THANK YOU

Helle Lauridsen