Mollusks and their shell biomineralization

Mollusks (120 000 living species)

Bivalves
(12 000 species)

Solenogasters
(350 species)

Caudofoveates
(250 species)

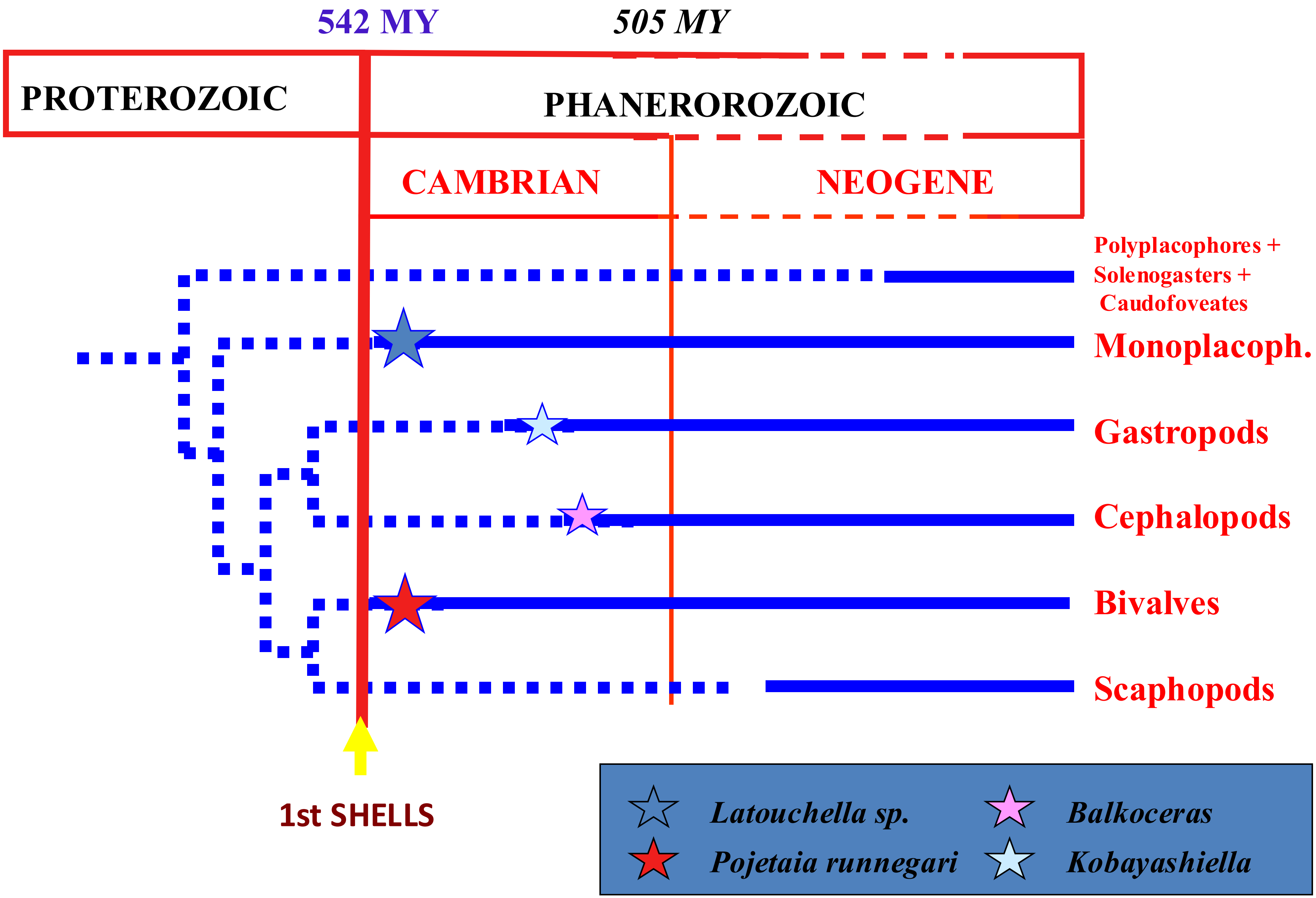
Scaphopods
(400 species)

Polyplacophores
(900 species)

Gastropods
(103 000 species)

Cephalopods
(700 species)

Monoplacophorans
(15 species)



542 MY

505 MY

PROTEROZOIC

PHANEROROZOIC

CAMBRIAN

NEOGENE

Polyplacophores +
Solenogasters +
Caudofoveates

Monoplacoph.

Gastropods

Cephalopods

Bivalves

Scaphopods

1st SHELLS

★	***Latouchella sp.***	★	***Balkoceras***
★	***Pojetaia runnegari***	☆	***Kobayashiella***

- ***Since the Cambrian, continuous fossil record***
- ***Great adaptability***
- ***Evolutionary success: efficient shelter against predation and dessiccation***

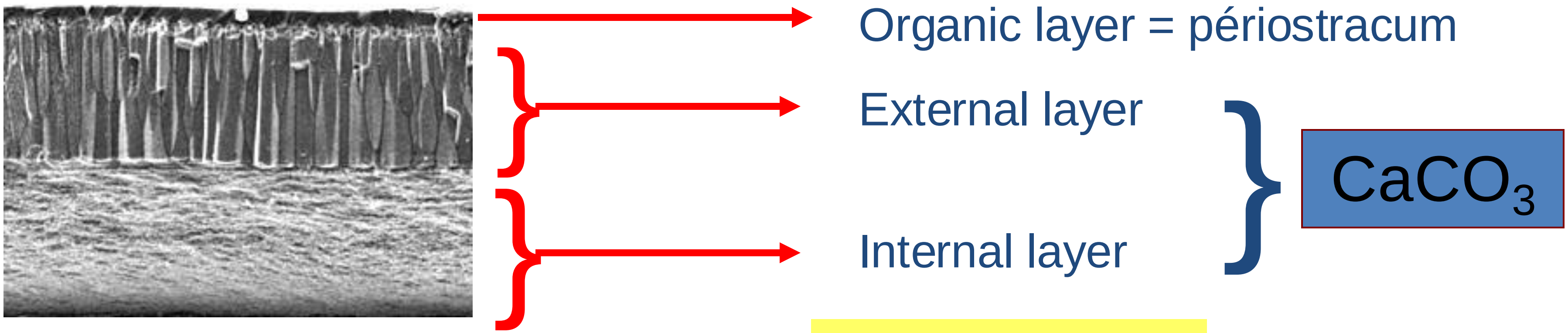
One of the keys of the

evolutionary success:

THE SHELL

The shell, a multi-layered material


Servaco-Géol.UB© 2005.JT


Organic layer = périostracum

External layer

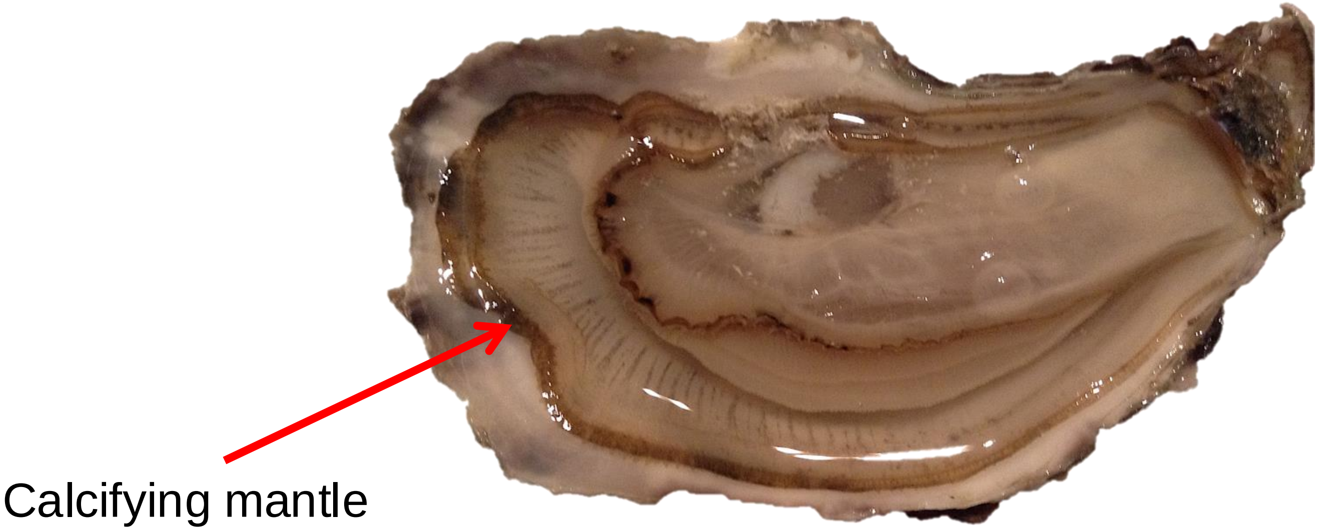
Internal layer

$CaCO_3$

	SHELL
Calcite	YES
Aragonite	YES
Vatérite	Traces
Monohydrocalcite	0
Protodolomite	0
Amorphe (ACC)	Traces

Physiology of calcification

Crassostrea gigas

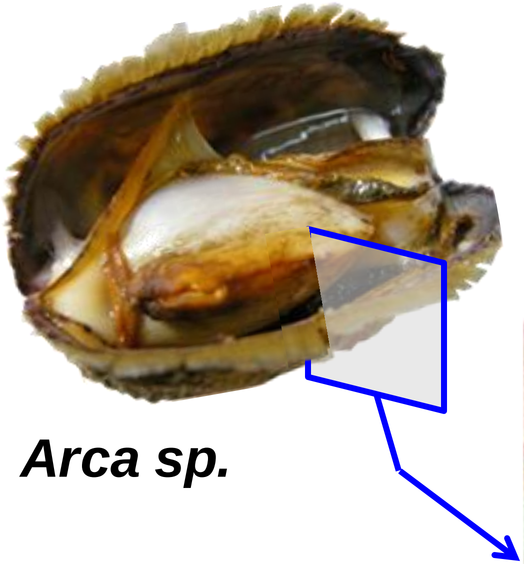


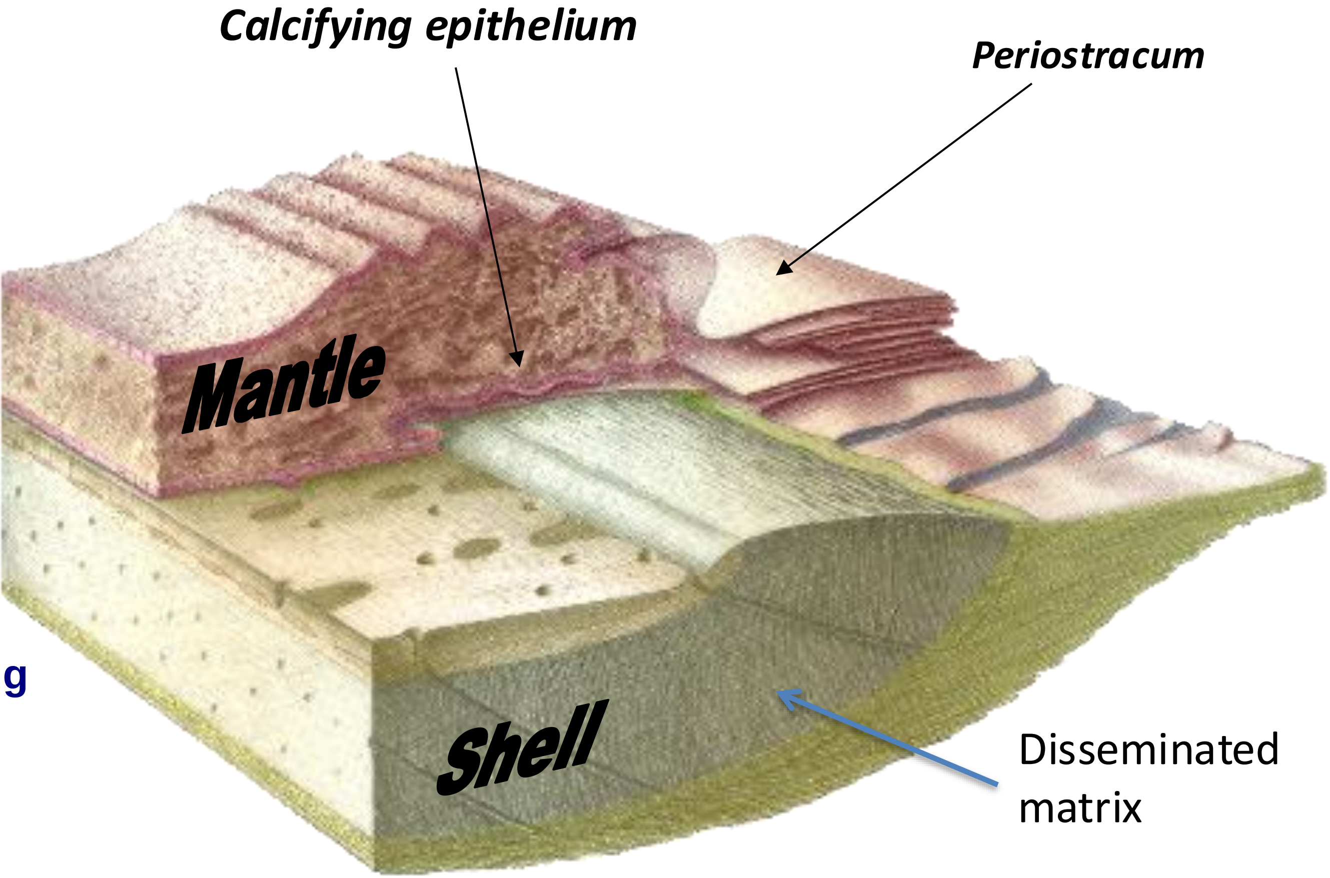
$$HCO_3^- + Ca^{2+} \longrightarrow CaCO_3 + H^+$$

- Translocation of precursor ions & extrusion
- Secretion of amorphous granules
- Secretion of skeletal matrix
- Proton reabsorption

The mantle–shell interface in mollusks

http://tiritinyam.blogspot.com/2007/12/les-caixetes-de-vinars.html



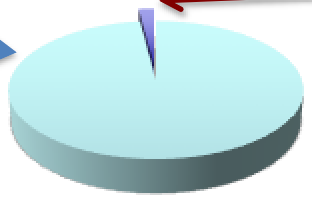


From Marin et al., 2012,
Adapted from Waller, 1980

Interface mantle / growing shell: the ‘extrapallial space’:

- **Inorganic precursors**
- **Matrix**
- **‘Self-assembling process’**

The shell, an organo-mineral material…



$CaCO_3$
+ Minor elements:
Mg, Sr…

Organic matrix

0,05 à 1-2 % of the shell

PROTEINS & PEPTIDES | POLYSACCHARIDES | LIPIDS | PIGMENTS | METABOLITES

Well studied → **Poorly studied**

Mollusk shell matrix

- **Displays major roles in biomineralization**

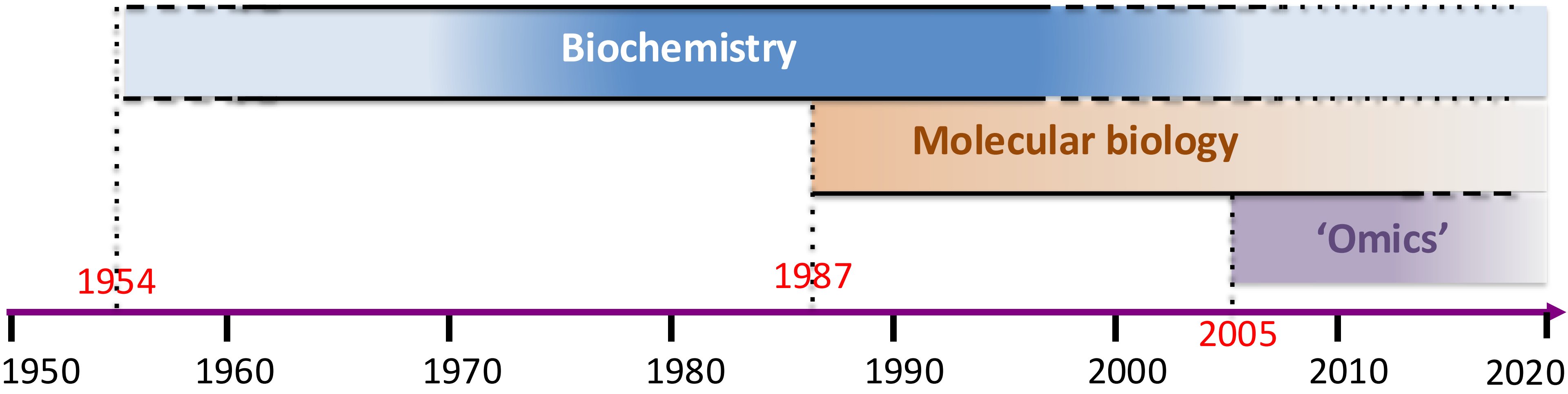
« MOLECULAR TOOLBOX ! »

& the main regulator of mineral deposition

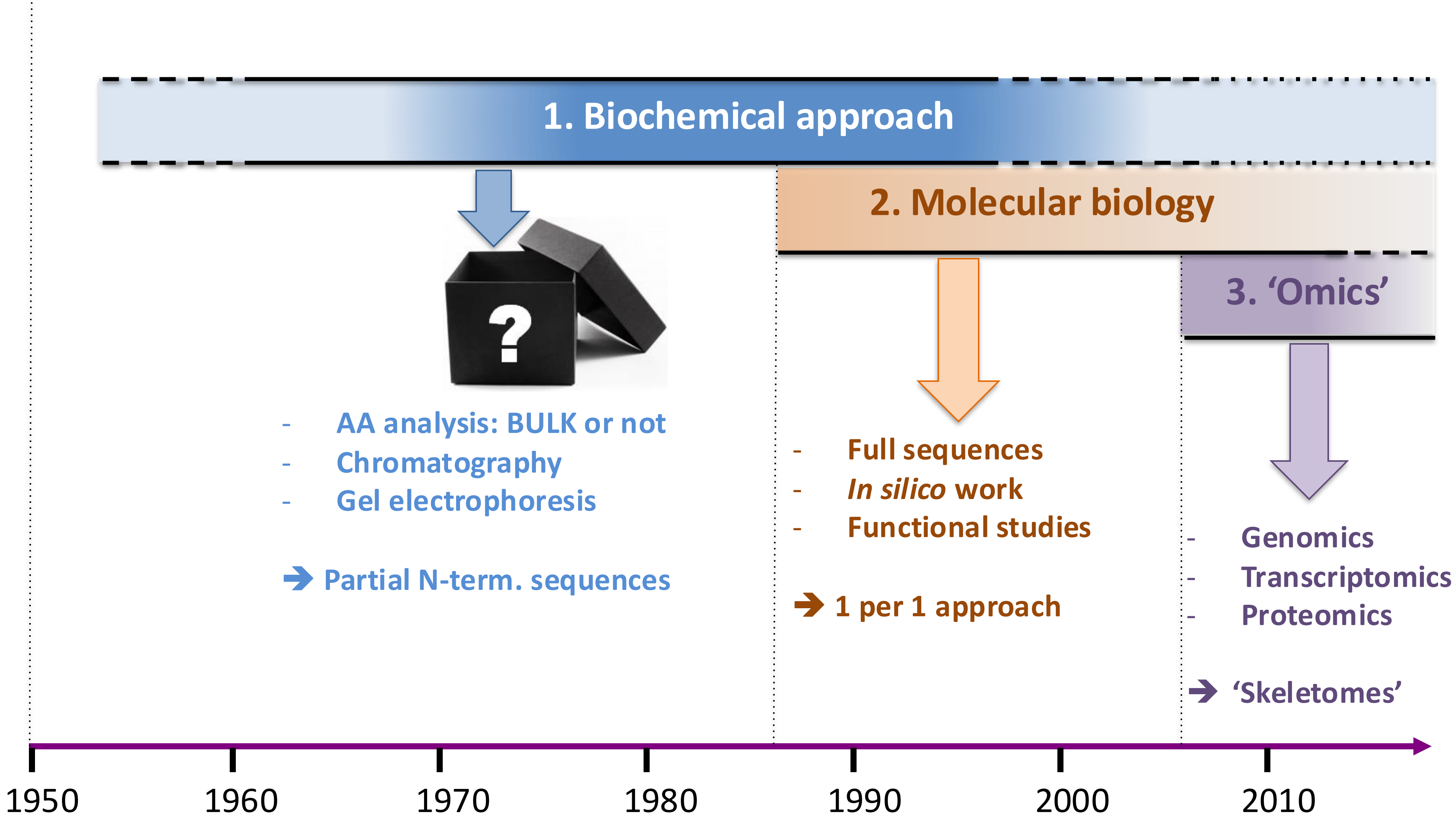
Drastic evolution of the concept of « skeletal matrix » in the last decades…

The manner we perceive the shell matrix and its putative functions

has a lot to do with the techniques we use to analyze it !!



Mollusk shell matrix



1. Biochemical approach

- AA analysis: BULK or not
- Chromatography
- Gel electrophoresis

➔ Partial N-term. sequences

2. Molecular biology

- Full sequences
- *In silico* work
- Functional studies

➔ 1 per 1 approach

3. 'Omics'

- Genomics
- Transcriptomics
- Proteomics

➔ 'Skeletomes'

(From Marin *et al.*, 2016)

Mollusk shell matrix

1. Biochemical approach

- AA analysis: BULK or not
- Chromatography
- Gel electrophoresis

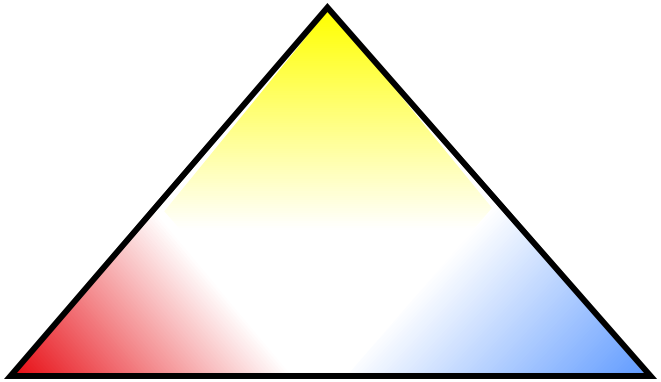
➔ Partial N-term. sequences

- Proteins = molds around crystals or act as physical templates (hydrophobic)
- Proteins that bind calcium ions (polyanionic, Asp-rich)

SIMPLE MOLECULAR MODELS

IN THE MID-EIGHTIES: *aa analysis, chromatography... No sequence data !!*

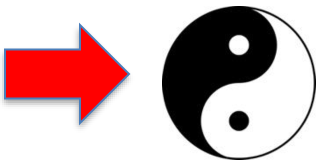
The ternary
« sandwich model »
for mollusks (nacre)



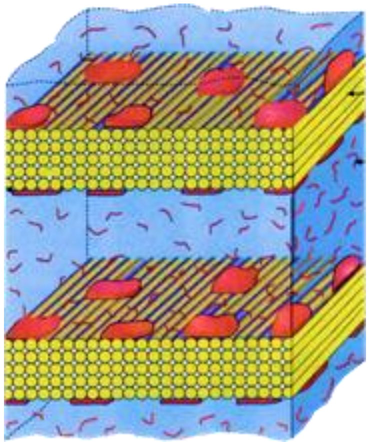
Chitin fibers
STRUCTURAL SUPPORT

Soluble acidic proteins
NUCLEATION
& INHIBITION

Insoluble hydrophobic proteins
MOLD, TEMPLATE
then HYDROGEL

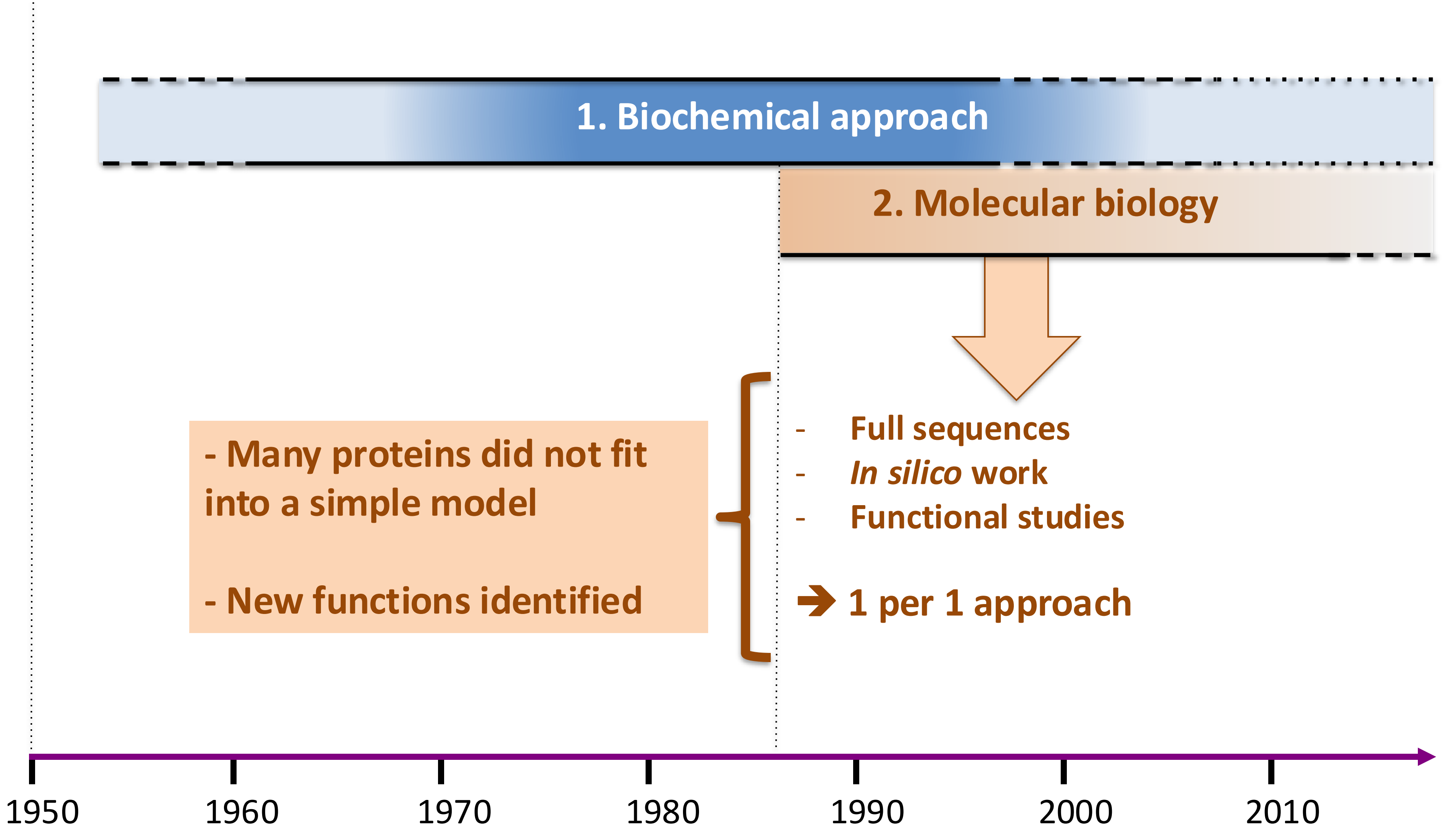


Yin & Yang



Levi-Kalisman
et al., 2001

Mollusk shell matrix



1. Biochemical approach

2. Molecular biology

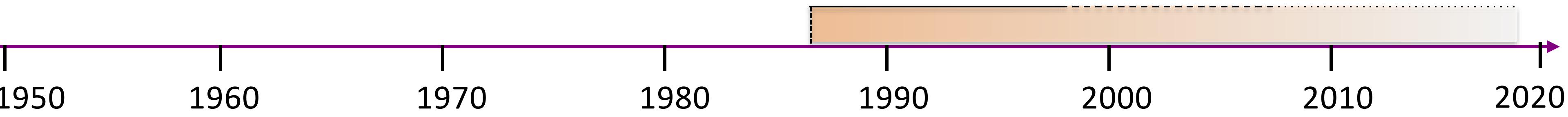
- Full sequences
- *In silico* work
- Functional studies

➔ 1 per 1 approach

- Many proteins did not fit into a simple model

- New functions identified

Analysis of the shell proteins: the molecular biology approach

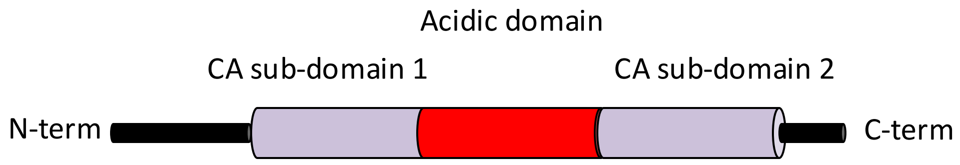


- Full sequences of shell matrix proteins
- Between 1996 and 2008, > 40 proteins for mollusks
- Overexpression and functional assays.
- **1 per 1 approach**.

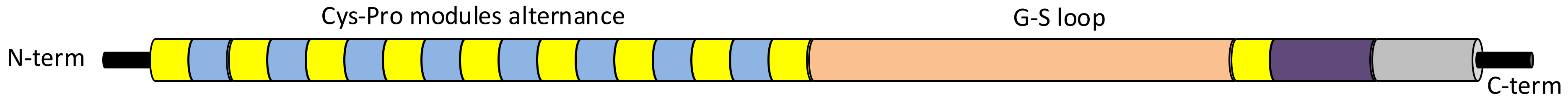
Most of the identified proteins did not fit into a ternary model

Examples:

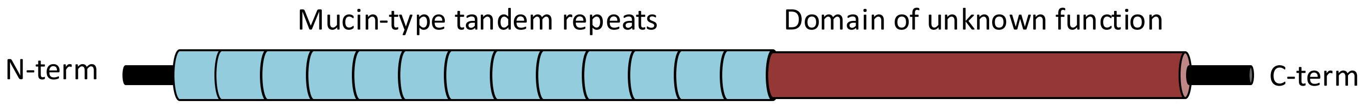
Nacrein *(Miyamoto et al., 1996)*



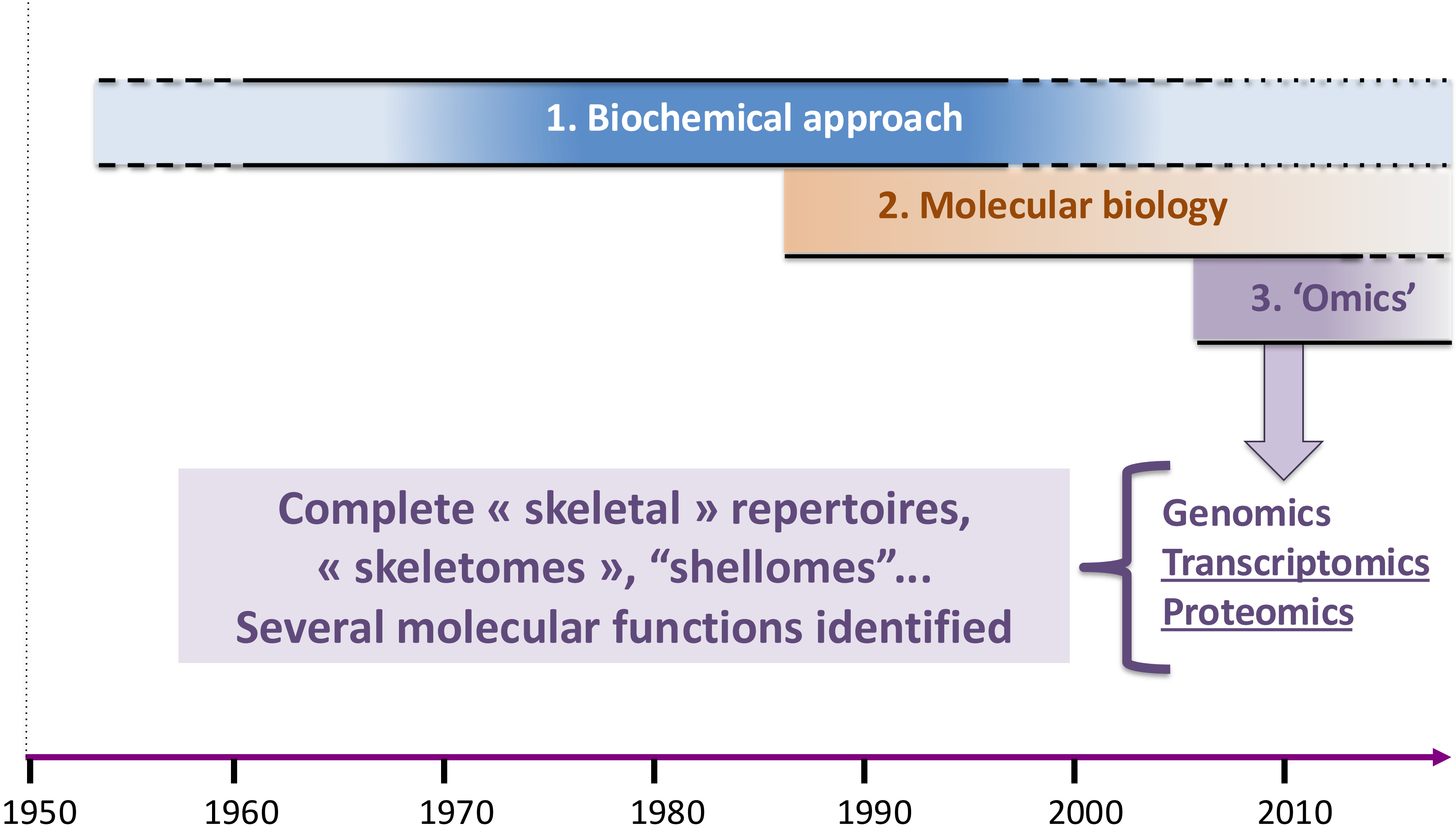
Lustrin A *(Shen et al., 1997)*



Mucoperlin *(Marin et al., 2000)*



Mollusk shell matrix



1. Biochemical approach

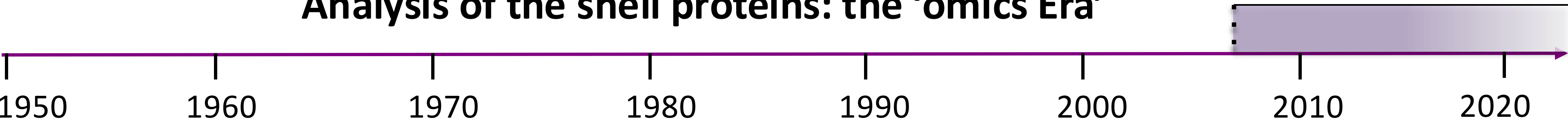
2. Molecular biology

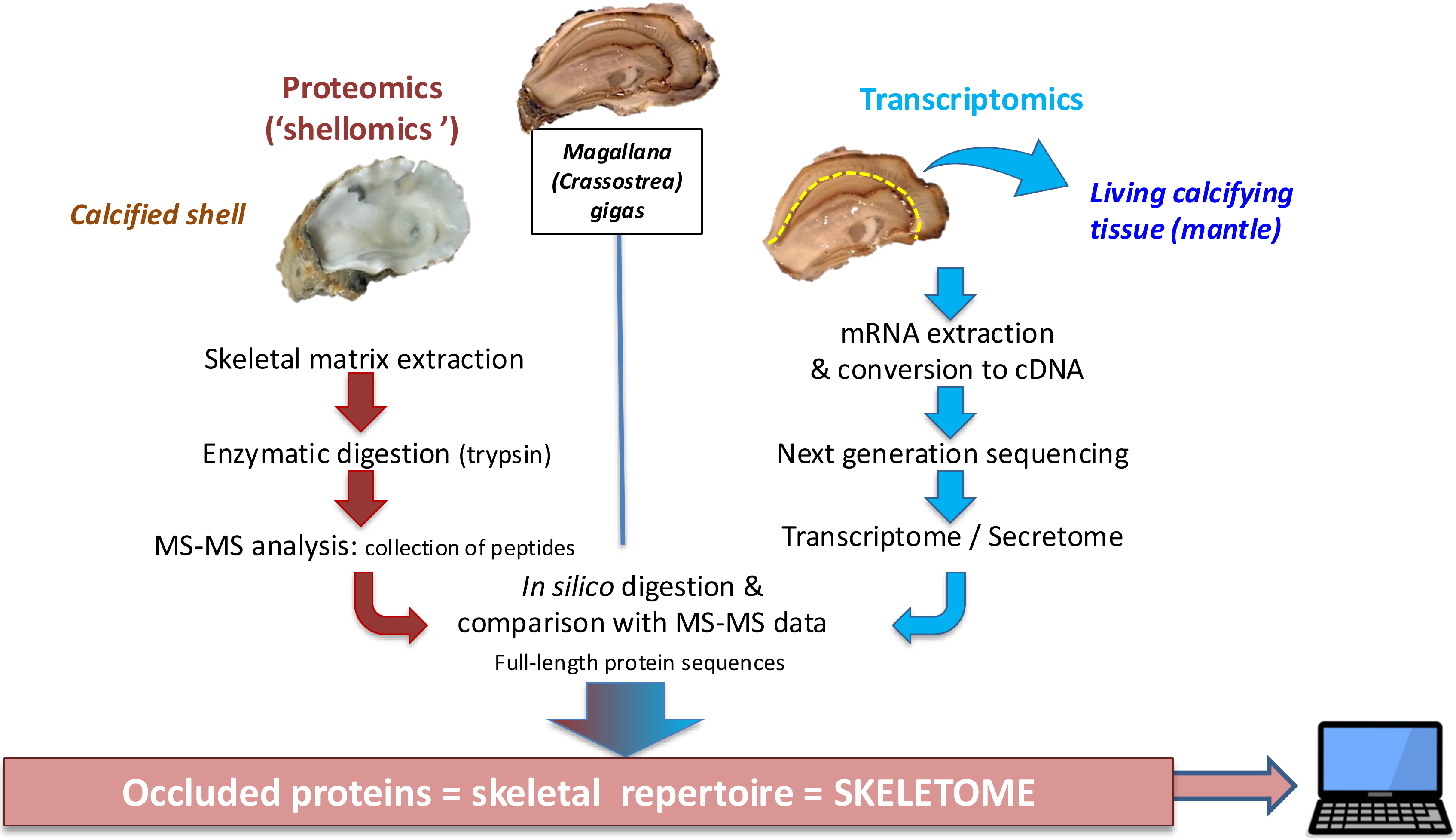
3. ‘Omics’

Complete « skeletal » repertoires,
« skeletomes », “shellomes”...
Several molecular functions identified

Genomics
Transcriptomics
Proteomics

Analysis of the shell proteins: the ‘omics Era’





Proteomics (‘shellomics ’)

Calcified shell

Skeletal matrix extraction

Enzymatic digestion (trypsin)

MS-MS analysis: collection of peptides

Magallana (Crassostrea) gigas

Transcriptomics

Living calcifying tissue (mantle)

mRNA extraction
& conversion to cDNA

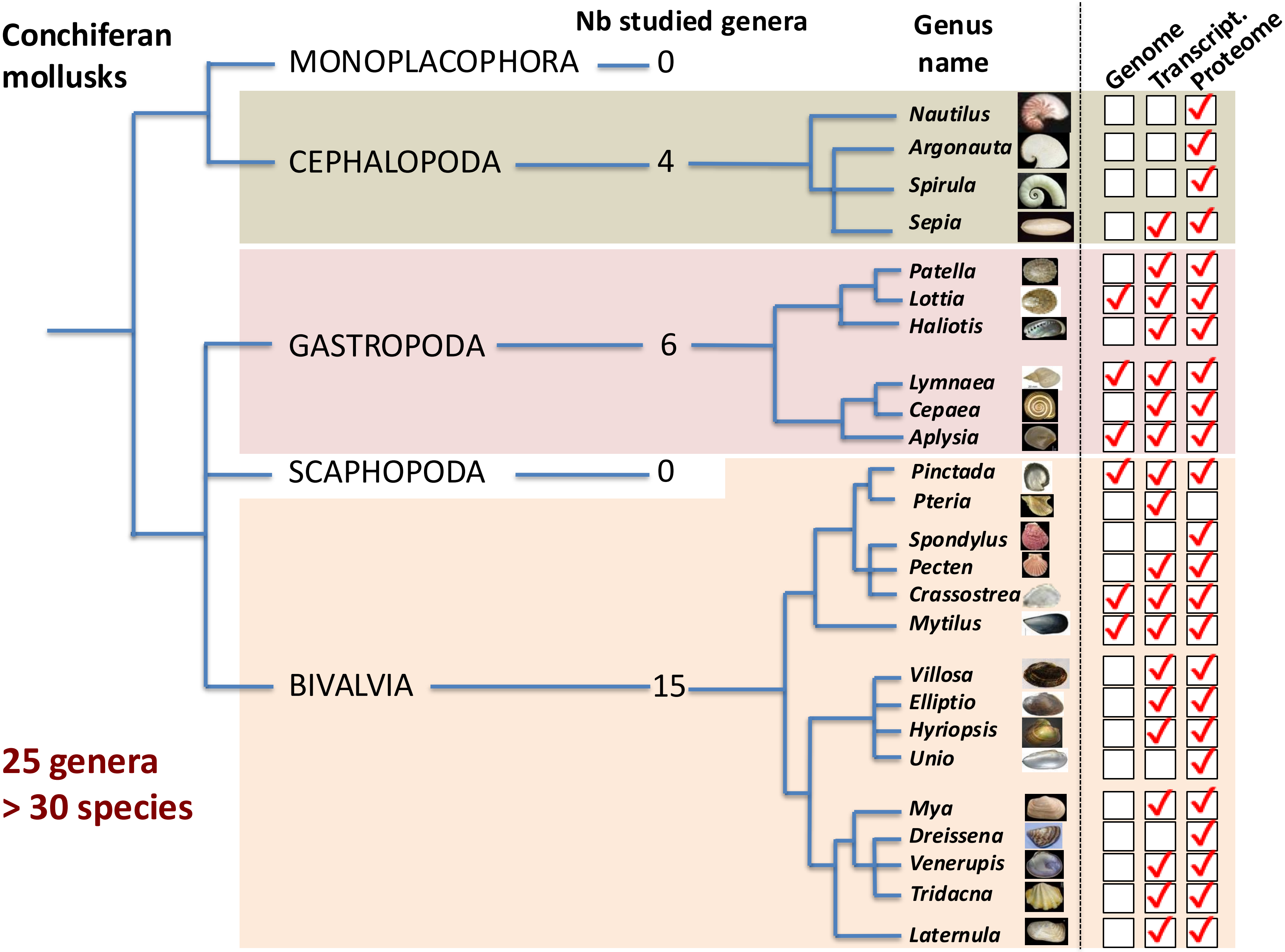
Next generation sequencing

Transcriptome / Secretome

In silico digestion &
comparison with MS-MS data

Full-length protein sequences

Occluded proteins = skeletal repertoire = SKELETOME



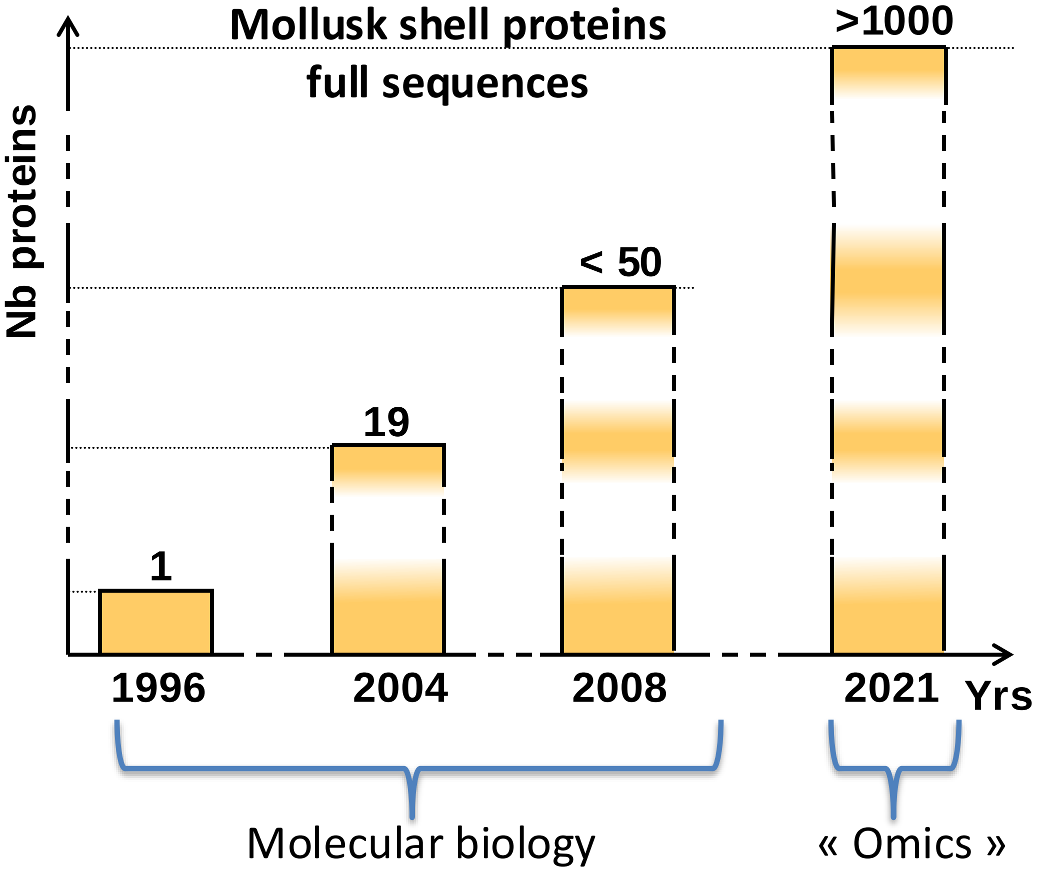
Conchiferan mollusks	Nb studied genera
MONOPLACOPHORA	0
CEPHALOPODA	4
GASTROPODA	6
SCAPHOPODA	0
BIVALVIA	15

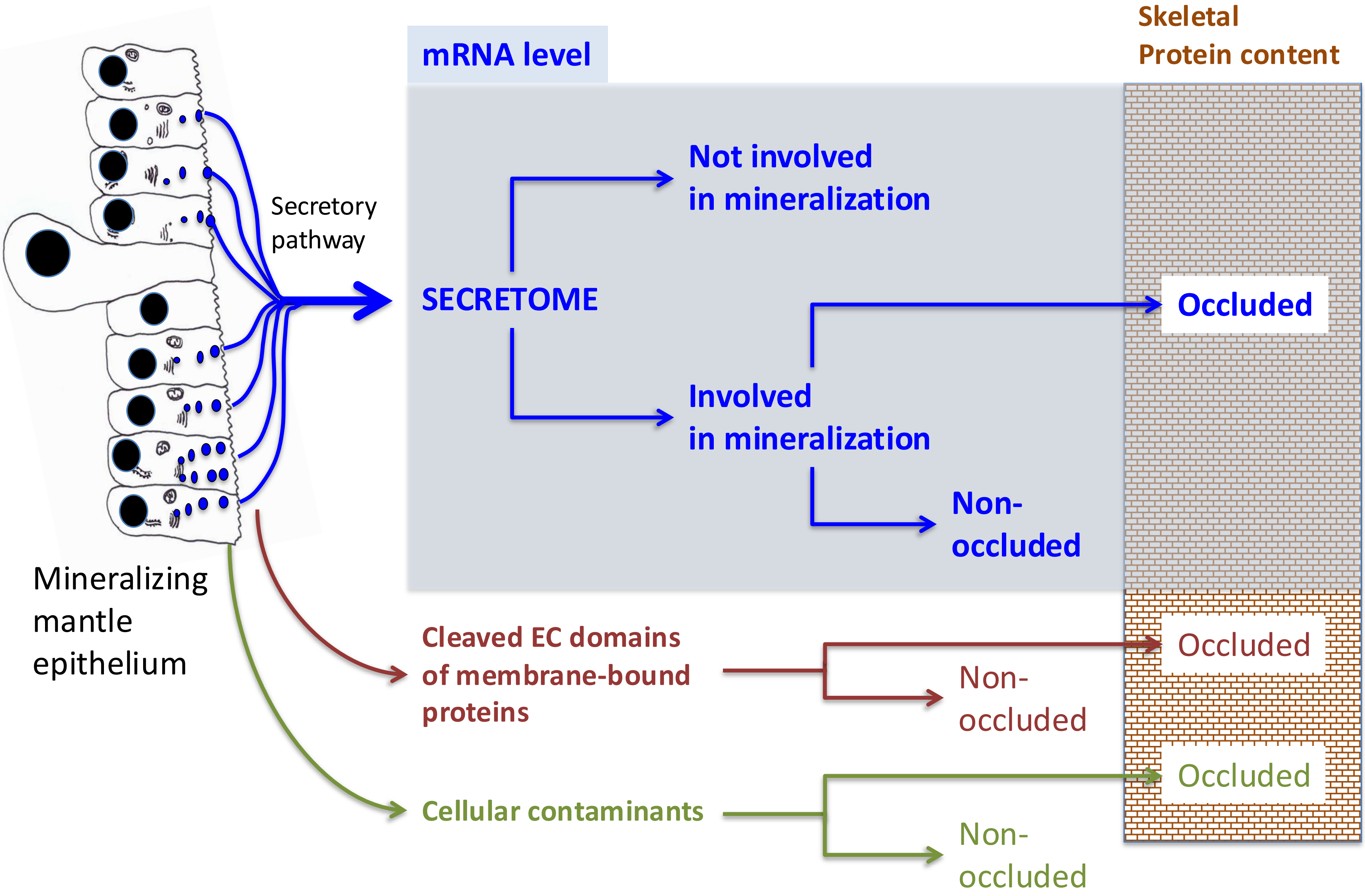
Genus name	Genome	Transcript.	Proteome
Nautilus	☐	☐	☑
Argonauta	☐	☐	☑
Spirula	☐	☐	☑
Sepia	☐	☑	☑
Patella	☐	☑	☑
Lottia	☑	☑	☑
Haliotis	☐	☑	☑
Lymnaea	☑	☑	☑
Cepaea	☐	☑	☑
Aplysia	☑	☑	☑
Pinctada	☑	☑	☑
Pteria	☐	☑	☐
Spondylus	☐	☐	☑
Pecten	☐	☑	☑
Crassostrea	☑	☑	☑
Mytilus	☑	☑	☑
Villosa	☐	☑	☑
Elliptio	☐	☑	☑
Hyriopsis	☐	☑	☑
Unio	☐	☐	☑
Mya	☐	☑	☑
Dreissena	☐	☐	☑
Venerupis	☐	☑	☑
Tridacna	☐	☑	☑
Laternula	☐	☑	☑

25 genera
> 30 species

Contribution of the 'omics' techniques

- Massive increase of the number of identified proteins
- Concept of skeletal matrix: blurred
- Several new functions identified
- Majority of novel unknown functions
- Are skeletal proteins all involved in mineralization ?
  ➔ cytoplasmic contaminants
  Necessity to clean properly skeletal tissues





Skeletal Protein content

mRNA level

Secretory pathway

SECRETOME

Not involved in mineralization

Involved in mineralization

Occluded

Non-occluded

Mineralizing mantle epithelium

Cleaved EC domains of membrane-bound proteins

Occluded

Non-occluded

Cellular contaminants

Occluded

Non-occluded

The facts

- >> 1000 proteins
- Several functional domains
- Several protein families
- Low Complexity Domains

THE PROBLEMS

- Concept / outlines of skeletal matrix: blurred
- How to relate protein sequences and functions ?
- How do all these proteins work together ?

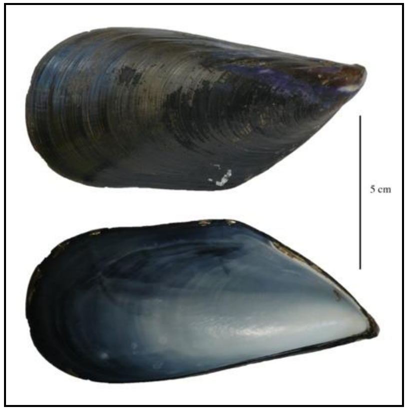
The possibilities

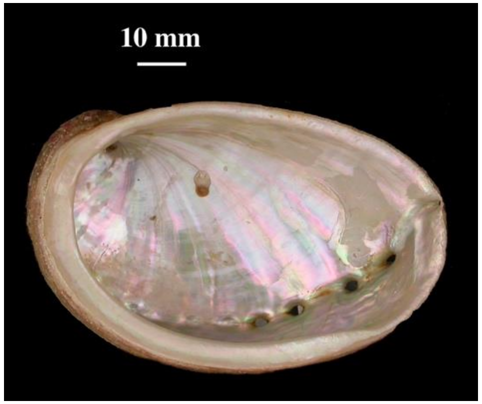
- Comparison between calcifying repertoires
- Sketch macroevolutionary trends
- Diagenesis of skeletal matrix

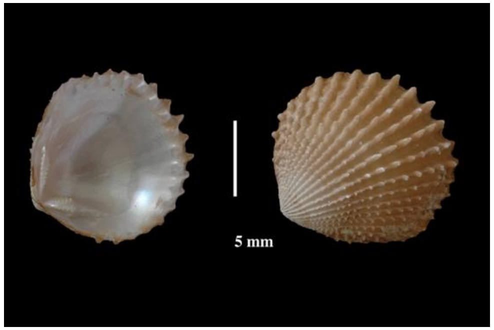
Example 1: *nacre in molluscs, single or multiple inventions ?*

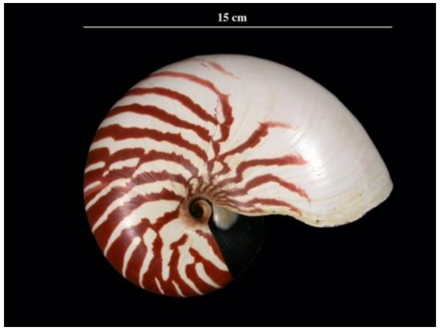
* ***One of the several shell microstructures***
* ***Always internal layer***
* ***Laminar microstructure***

* ***Nacre = typical of molluscs***
* ***Always aragonitic***
* ***Cambrian origin***





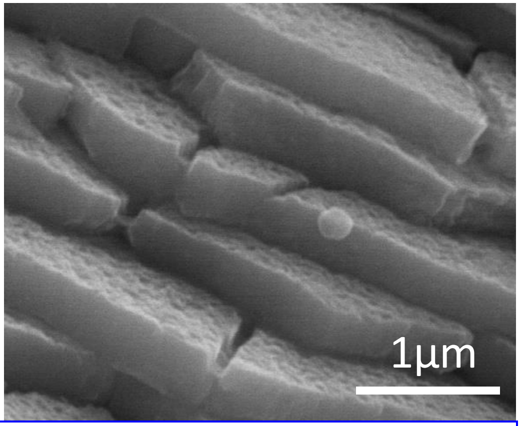


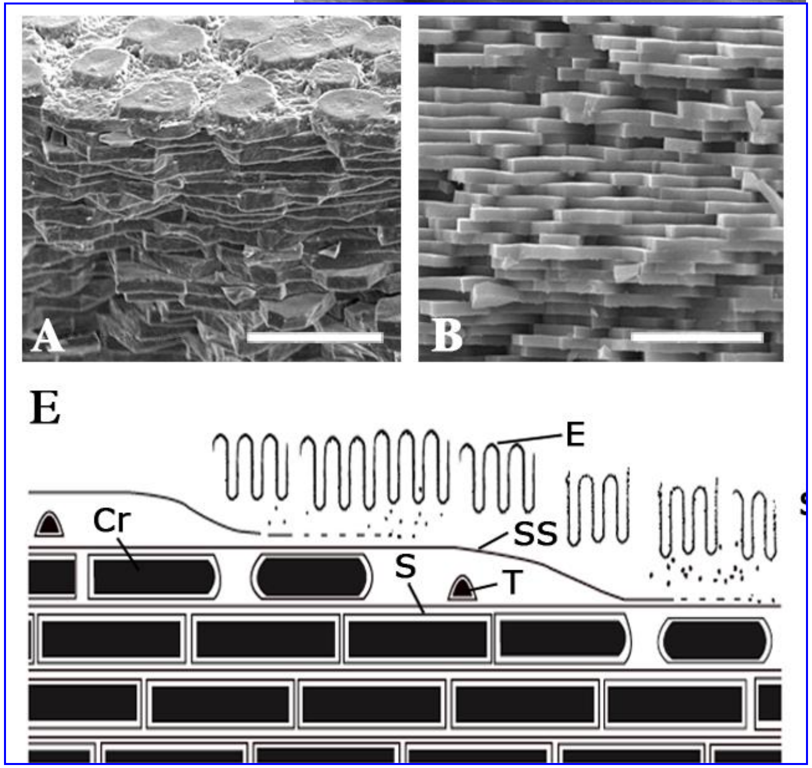


Nacre in mollusks

Brickwall nacre

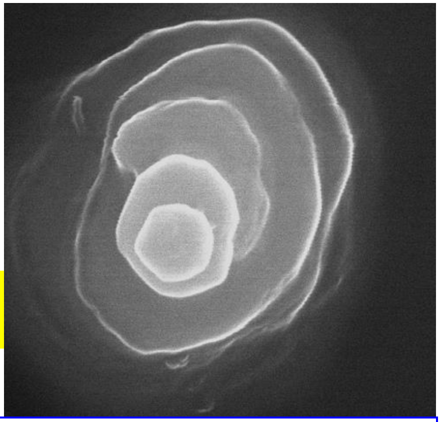
BIVALVES

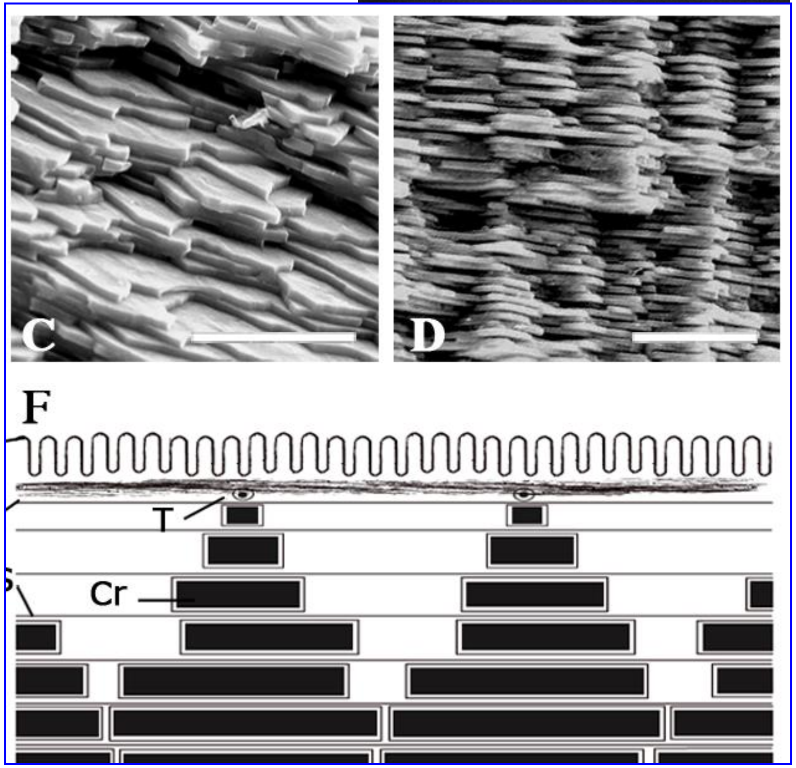




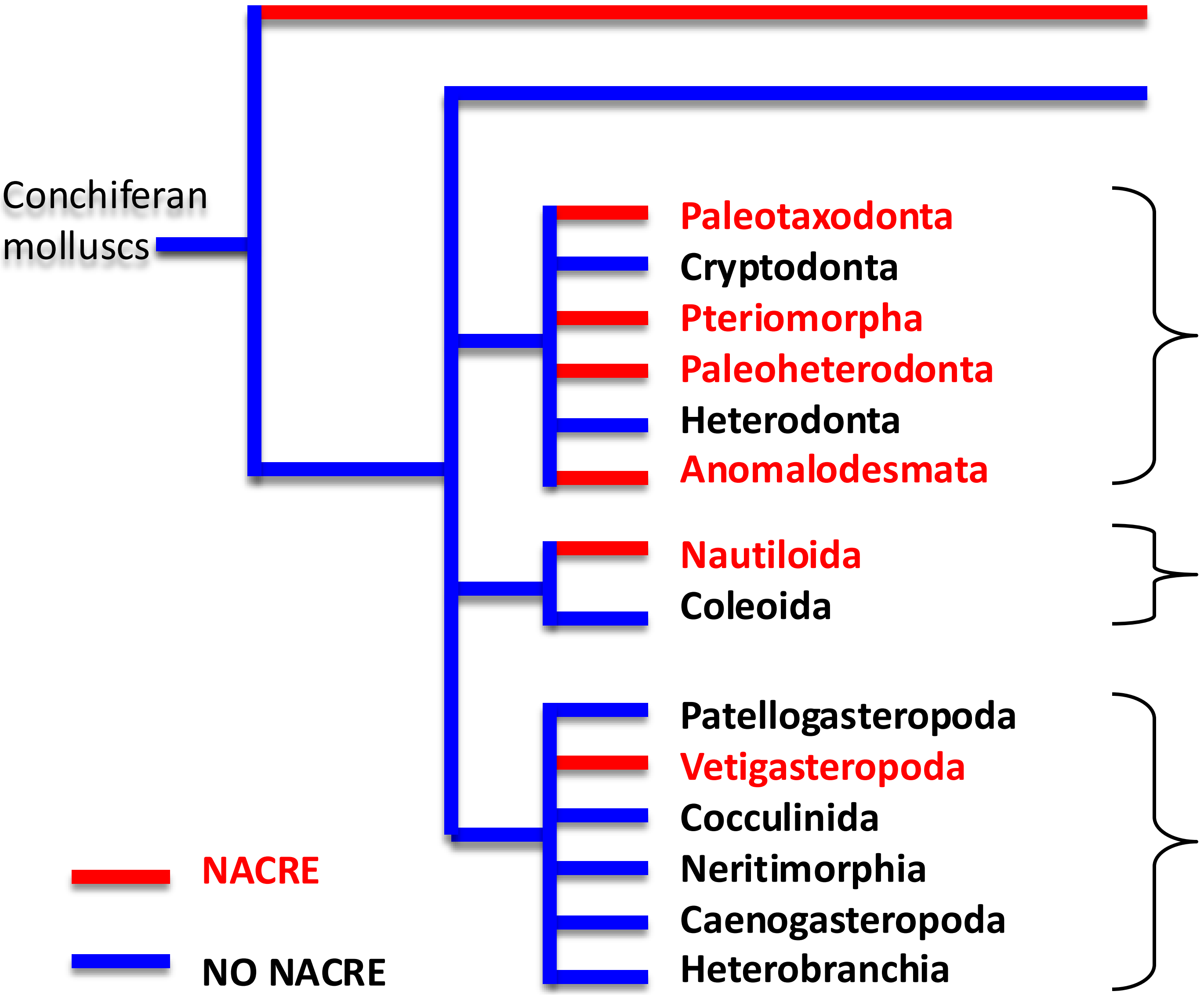
Columnar nacre

GASTROPODS





Example 1



Conchiferan molluscs

Monoplacophorans

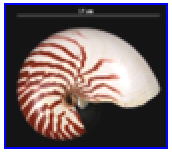
Scaphopods

Paleotaxodonta
Cryptodonta
Pteriomorpha
Paleoheterodonta
Heterodonta
Anomalodesmata

Bivalves

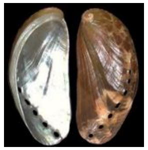
Nautiloida
Coleoida

Cephalopods



Patellogasteropoda
Vetigasteropoda
Cocculinida
Neritimorphia
Caenogasteropoda
Heterobranchia

Gastropods



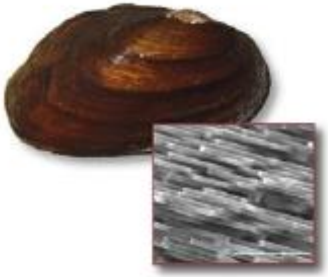
🟥 NACRE

🟦 NO NACRE

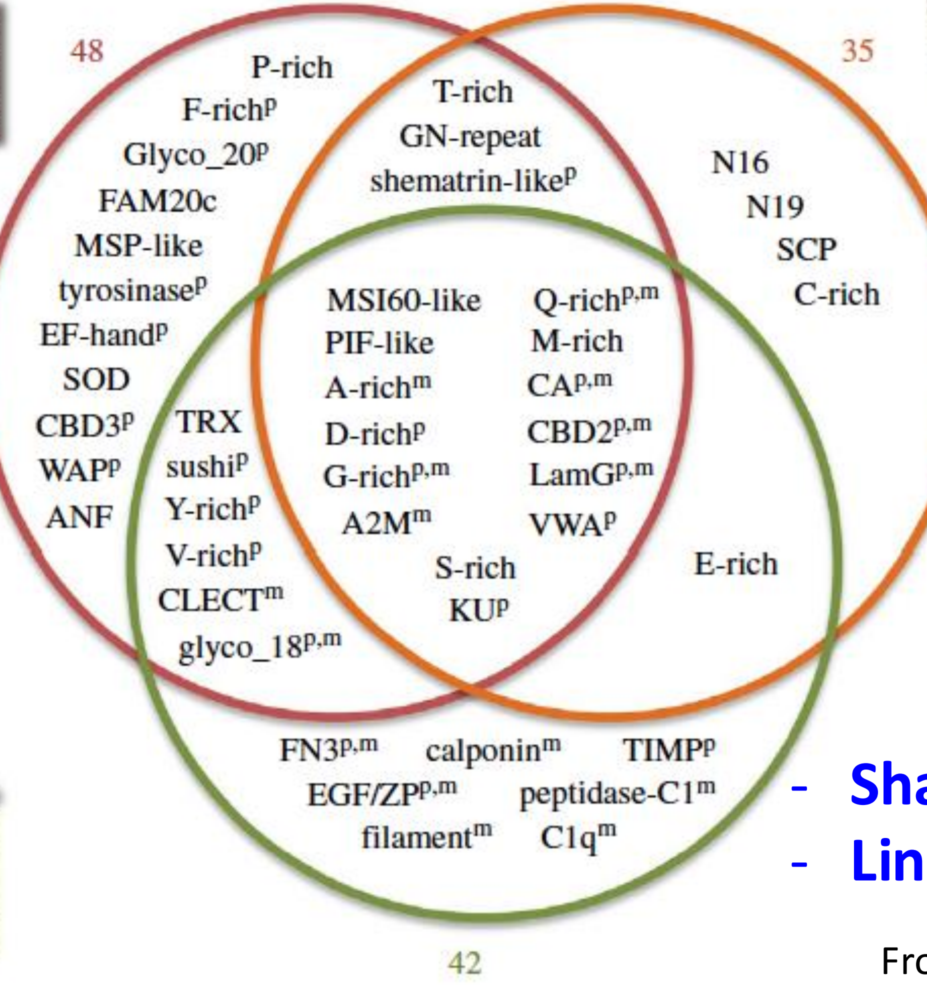
Example 1

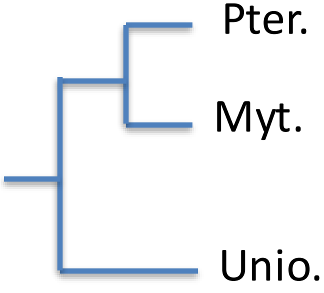
Within nacreous bivalves…

UNIONOIDA



PTERIOIDA





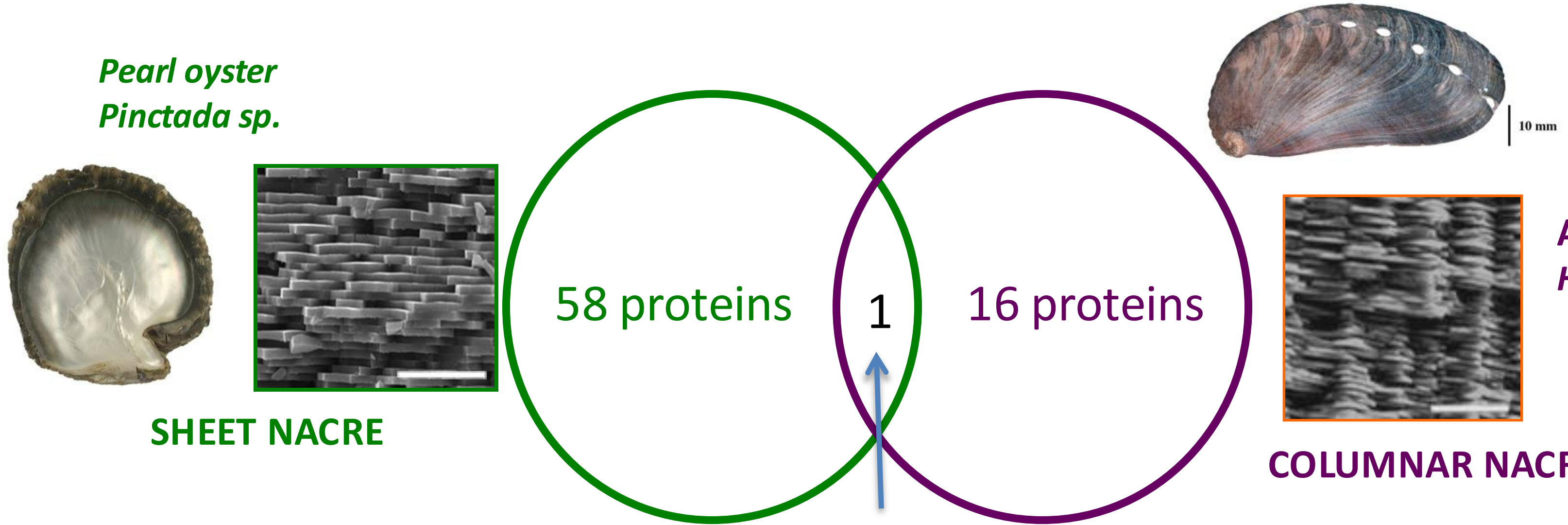
MYTILOIDA

- **Shared**
- **Lineage-specific**

From Marie *et al.*, 2017

Example 1

Nacreous bivalve *vs.* nacreous gastropod



Pearl oyster
Pinctada sp.

SHEET NACRE

58 proteins

1

16 proteins

Kunitz-like protein

Abalone
H. asinina

COLUMNAR NACRE

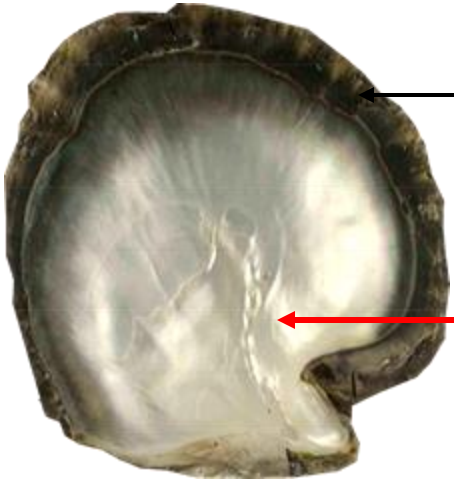
Adapted from Marie *et al.*, 2012

- **Within bivalves, <u>several similarities</u>, but also <u>taxon-specific nacre proteins</u>**
- **Between bivalves & gastropods, very different nacre proteins assemblages**

- **=> Independent inventions…, or fast evolution of nacre proteins ("drift" from a common nacre ancestor) ?**

- **Different assemblages can build similar microstructures: plasticity of the system => which 'driving force' constrains the system to produce nacre?**

Example 2: *prisms and nacre, same matrices ?*



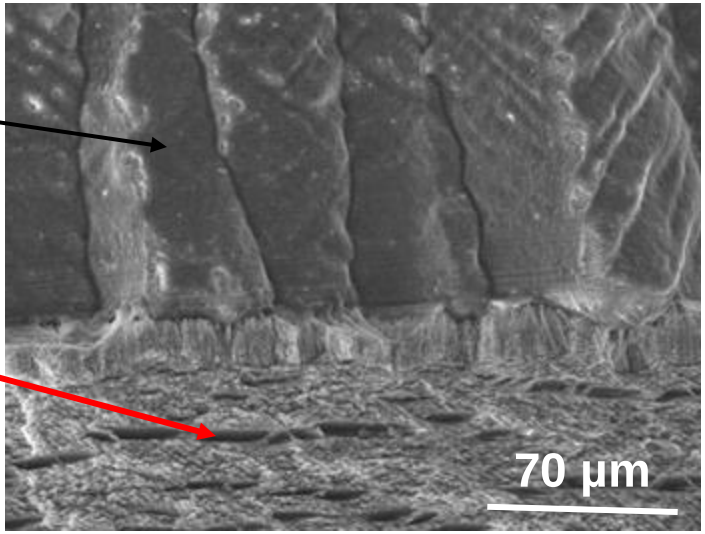
Pinctada margaritifera

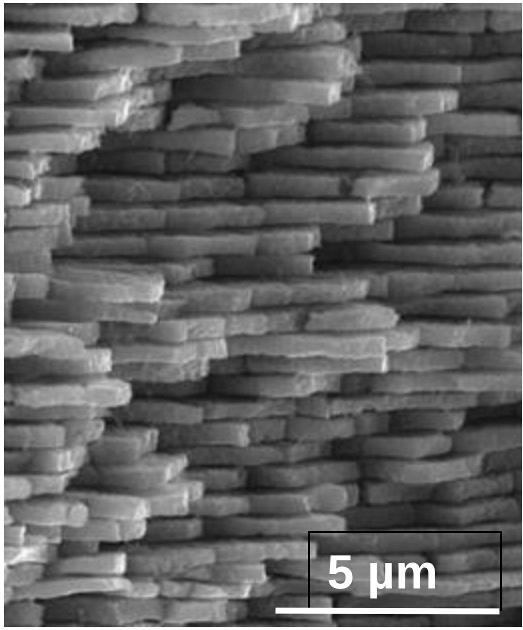
Prisms

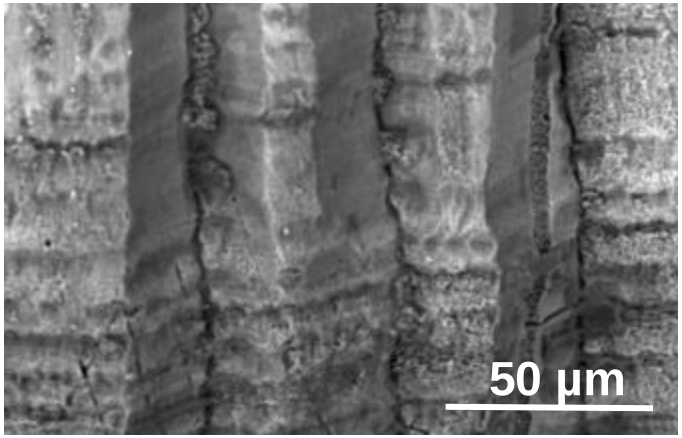
(calcite)

Nacre

(aragonite)





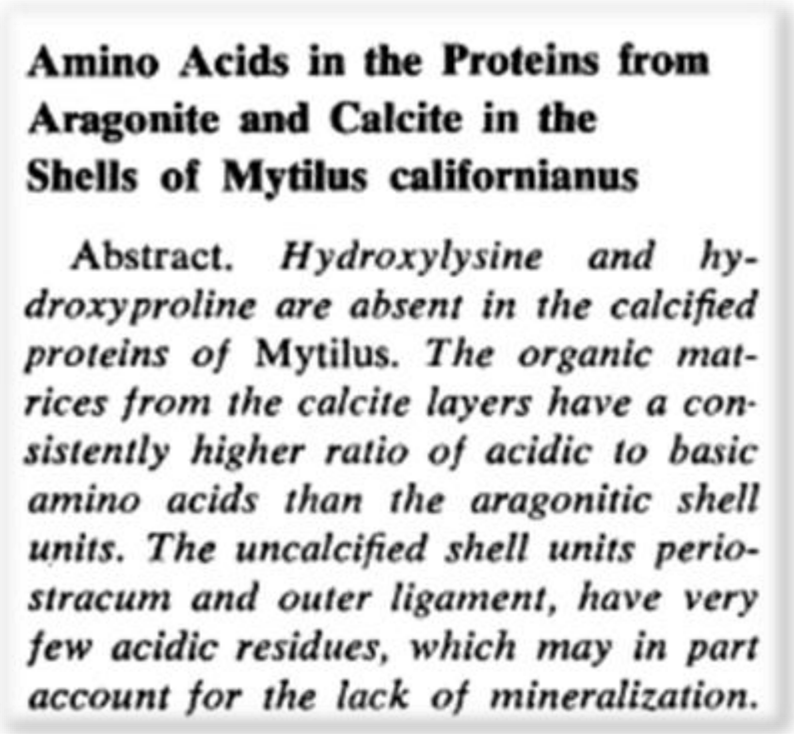


ADEQUA Consortium

Are prisms and nacre made from the same shell matrix repertoire ?

Example 2

A fundamental question that was left open for decades...



Amino Acids in the Proteins from Aragonite and Calcite in the Shells of Mytilus californianus

Abstract. *Hydroxylysine and hydroxyproline are absent in the calcified proteins of* Mytilus. *The organic matrices from the calcite layers have a consistently higher ratio of acidic to basic amino acids than the aragonitic shell units. The uncalcified shell units periostracum and outer ligament, have very few acidic residues, which may in part account for the lack of mineralization.*

Hare, PE, *Science*, 1963

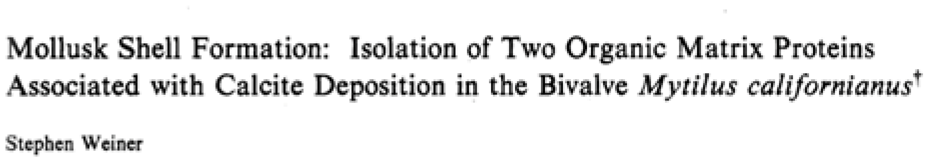
Biol. Rev. (1967), **42**, *pp.* 653–687 653

SUR LA STRUCTURE DES MATRICES ORGANIQUES DES COQUILLES DE MOLLUSQUES

PAR CH. GRÉGOIRE

Laboratoire de Biochimie, Université de Liège, Belgique

Grégoire, Ch, *Biol. Rev.,* 1967

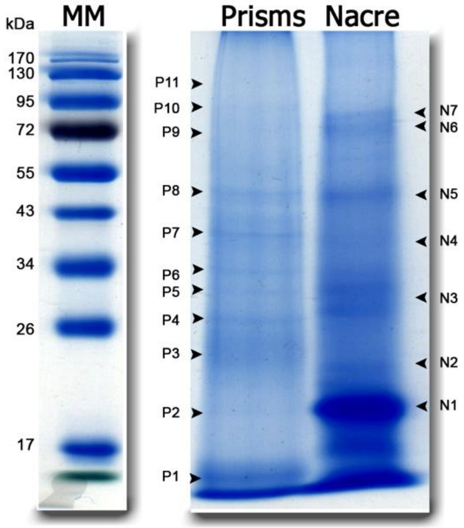


Mollusk Shell Formation: Isolation of Two Organic Matrix Proteins Associated with Calcite Deposition in the Bivalve *Mytilus californianus*†

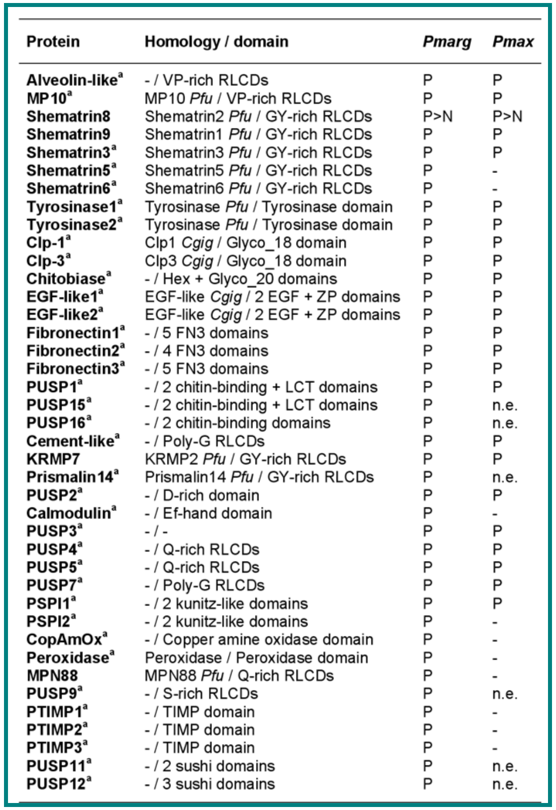
Stephen Weiner

Weiner, S, *Biochemistry*, 1983

Example 2

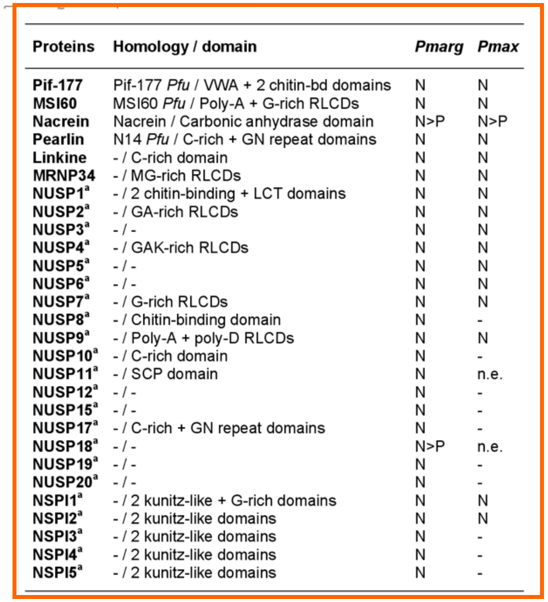


Prisms



Protein	Homology / domain	*Pmarg*	*Pmax*
Alveolin-like[a]	- / VP-rich RLCDs	P	P
MP10[a]	MP10 *Pfu* / VP-rich RLCDs	P	P
Shematrin8	Shematrin2 *Pfu* / GY-rich RLCDs	P>N	P>N
Shematrin9	Shematrin1 *Pfu* / GY-rich RLCDs	P	P
Shematrin3[a]	Shematrin3 *Pfu* / GY-rich RLCDs	P	P
Shematrin5[a]	Shematrin5 *Pfu* / GY-rich RLCDs	P	-
Shematrin6[a]	Shematrin6 *Pfu* / GY-rich RLCDs	P	-
Tyrosinase1[a]	Tyrosinase *Pfu* / Tyrosinase domain	P	P
Tyrosinase2[a]	Tyrosinase *Pfu* / Tyrosinase domain	P	P
Clp-1[a]	Clp1 *Cgig* / Glyco_18 domain	P	P
Clp-3[a]	Clp3 *Cgig* / Glyco_18 domain	P	P
Chitobiase[a]	- / Hex + Glyco_20 domains	P	P
EGF-like1[a]	EGF-like *Cgig* / 2 EGF + ZP domains	P	P
EGF-like2[a]	EGF-like *Cgig* / 2 EGF + ZP domains	P	P
Fibronectin1[a]	- / 5 FN3 domains	P	P
Fibronectin2[a]	- / 4 FN3 domains	P	P
Fibronectin3[a]	- / 5 FN3 domains	P	P
PUSP1[a]	- / 2 chitin-binding + LCT domains	P	P
PUSP15[a]	- / 2 chitin-binding + LCT domains	P	n.e.
PUSP16[a]	- / 2 chitin-binding domains	P	n.e.
Cement-like[a]	- / Poly-G RLCDs	P	P
KRMP7	KRMP2 *Pfu* / GY-rich RLCDs	P	P
Prismalin14[a]	Prismalin14 *Pfu* / GY-rich RLCDs	P	n.e.
PUSP2[a]	- / D-rich domain	P	P
Calmodulin[a]	- / Ef-hand domain	P	-
PUSP3[a]	- / -	P	P
PUSP4[a]	- / Q-rich RLCDs	P	P
PUSP5[a]	- / Q-rich RLCDs	P	P
PUSP7[a]	- / Poly-G RLCDs	P	P
PSPI1[a]	- / 2 kunitz-like domains	P	P
PSPI2[a]	- / 2 kunitz-like domains	P	-
CopAmOx[a]	- / Copper amine oxidase domain	P	-
Peroxidase[a]	Peroxidase / Peroxidase domain	P	-
MPN88	MPN88 *Pfu* / Q-rich RLCDs	P	-
PUSP9[a]	- / S-rich RLCDs	P	n.e.
PTIMP1[a]	- / TIMP domain	P	-
PTIMP2[a]	- / TIMP domain	P	-
PTIMP3[a]	- / TIMP domain	P	-
PUSP11[a]	- / 2 sushi domains	P	n.e.
PUSP12[a]	- / 3 sushi domains	P	n.e.

Nacre



Proteins	Homology / domain	*Pmarg*	*Pmax*
Pif-177	Pif-177 *Pfu* / VWA + 2 chitin-bd domains	N	N
MSI60	MSI60 *Pfu* / Poly-A + G-rich RLCDs	N	N
Nacrein	Nacrein / Carbonic anhydrase domain	N>P	N>P
Pearlin	N14 *Pfu* / C-rich + GN repeat domains	N	N
Linkine	- / C-rich domain	N	N
MRNP34	- / MG-rich RLCDs	N	N
NUSP1[a]	- / 2 chitin-binding + LCT domains	N	N
NUSP2[a]	- / GA-rich RLCDs	N	N
NUSP3[a]	- / -	N	N
NUSP4[a]	- / GAK-rich RLCDs	N	N
NUSP5[a]	- / -	N	N
NUSP6[a]	- / -	N	N
NUSP7[a]	- / G-rich RLCDs	N	N
NUSP8[a]	- / Chitin-binding domain	N	-
NUSP9[a]	- / Poly-A + poly-D RLCDs	N	N
NUSP10[a]	- / C-rich domain	N	-
NUSP11[a]	- / SCP domain	N	n.e.
NUSP12[a]	- / -	N	-
NUSP15[a]	- / -	N	-
NUSP17[a]	- / C-rich + GN repeat domains	N	-
NUSP18[a]	- / -	N>P	n.e.
NUSP19[a]	- / -	N	-
NUSP20[a]	- / -	N	-
NSPI1[a]	- / 2 kunitz-like + G-rich domains	N	N
NSPI2[a]	- / 2 kunitz-like domains	N	N
NSPI3[a]	- / 2 kunitz-like domains	N	-
NSPI4[a]	- / 2 kunitz-like domains	N	-
NSPI5[a]	- / 2 kunitz-like domains	N	-

Protein compositions =
VERY différent !!!

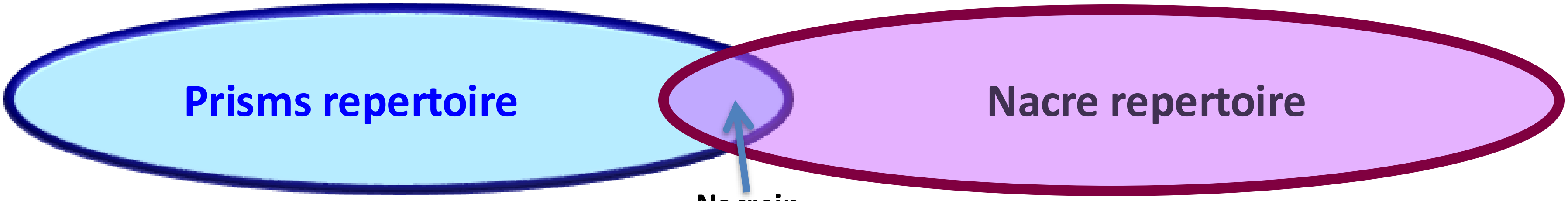
Identification *of potential biomarkers of shell calcification* **=>**

48 prisms proteins (40 new, 28 complete)

32 nacre proteins (26 new, 10 complete)

Example 2

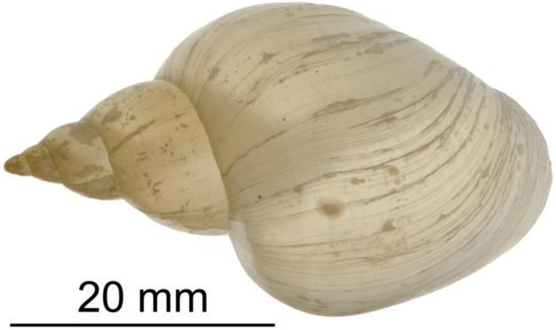
PRISMS	*Protein domains*	NACRE
Hydrophobic: VP-rich, Poly-G GY-rich Hydrophilic polar: Q-rich Acidic: D-rich Basic: K-rich	Low-Complexity (Repetitive or not)	Hydrophobic: G-rich, Poly-G, Poly-A, GN-rich, GA-rich, GAK-rich Sulfur-rich: C-rich, M-rich, Hydrophilic polar: N-rich Acidic: D-rich, E-rich, poly-D
Fibronectin EGF-like Sushi (adhesion/binding) Chitin-binding	ECM-like	SCP-like vWF Chitin-binding
Peroxidase Tyrosinase(phenol ox.) Cu-amine oxidase Clp (chitinase)	Enzyme	CA
TIMP (MMP inhibitor) Kunitz-like	Protease inhib.	Kunitz-like
Putative Ca-binding	EF-hand	



Nacrein
NUSP18
Shematrin8

Marie et al., PNAS, 2012

Example 3: mosaic composition of shell matrix proteins



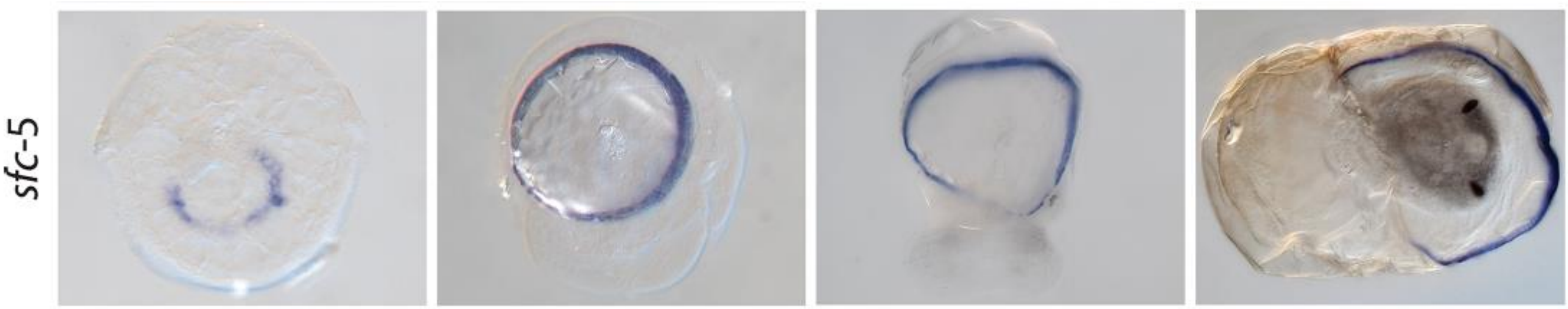
Lymnaea stagnalis

- Freshwater
- Pulmonate
- Belongs to a superfamily (Lymnaeoidea) that emerged in the Lower Jurassic (Dayrat et al., 2011)
- **Non nacreous: crossed-lamellar**

Herlitze et al., Gigascience, 2018
(Dan Jackson's group, Göttingen)

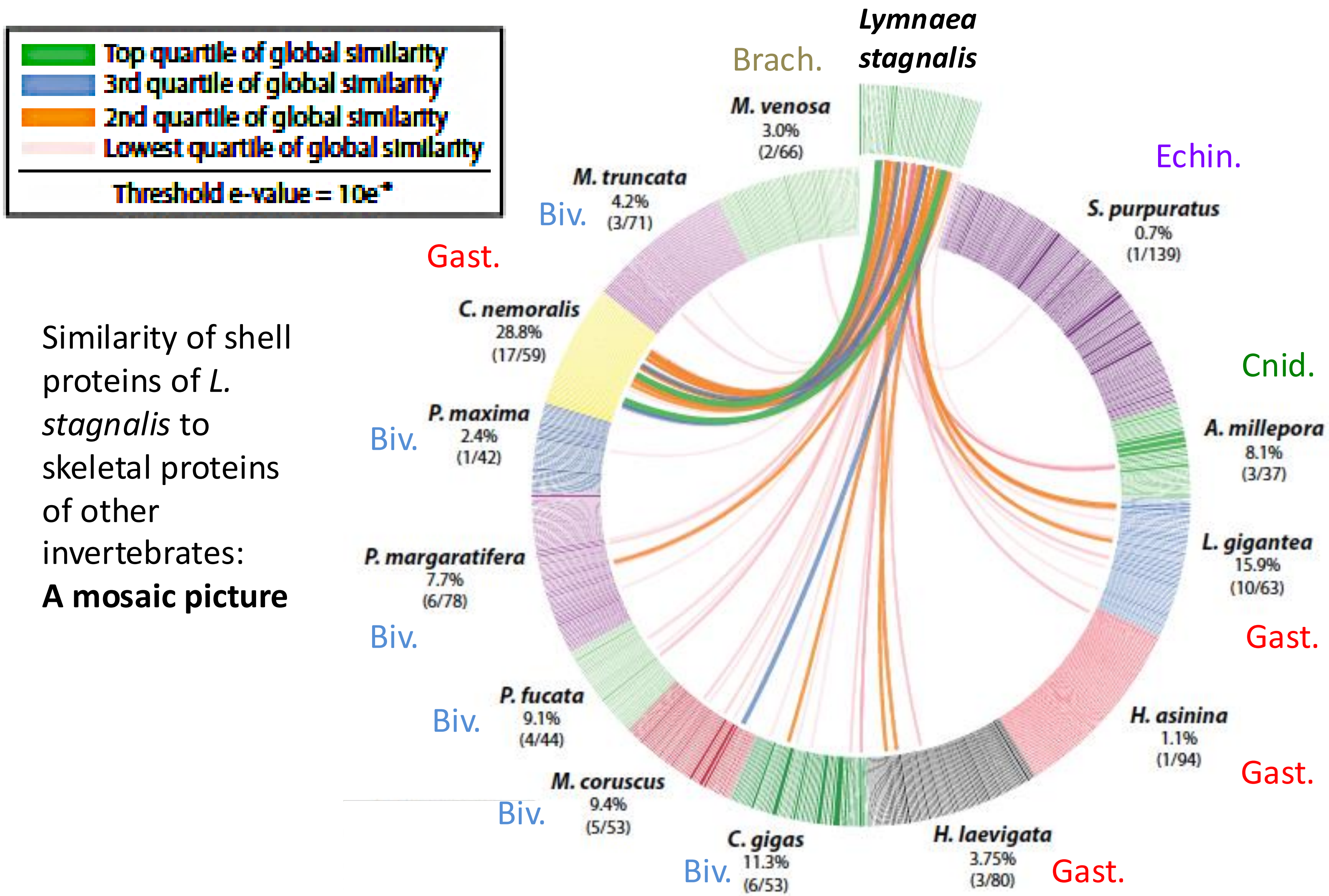
Proteomics of the shell matrix ➔ 40 proteins

For all of them, verification of the expression of the corresponding gene by *ISH* at different developmental stages



Example 3

Similarity of shell proteins of *L. stagnalis* to skeletal proteins of other invertebrates:
A mosaic picture



Herlitze et al., Gigascience, 2018

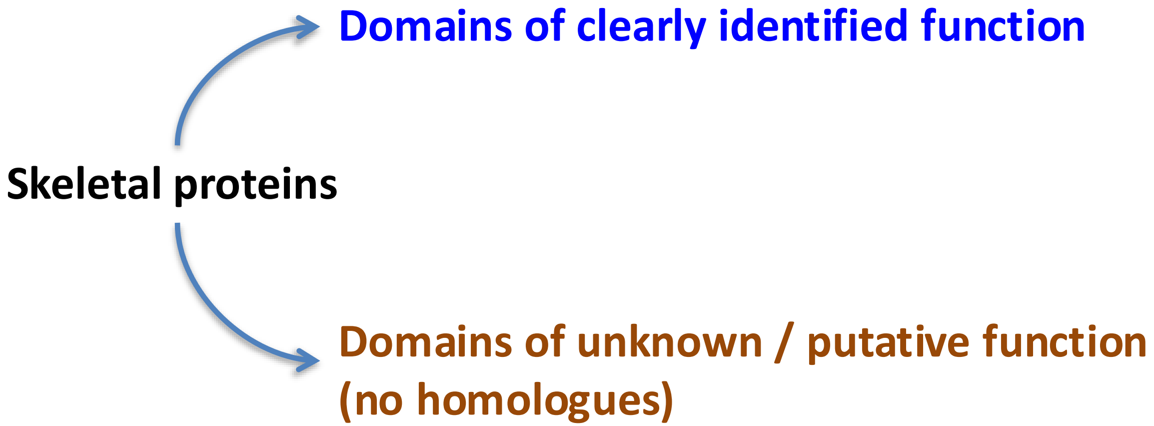
Skeletal (CaCO3) proteins, some characteristics…

* Two dozens to >> 100 per model **(depending on cleaning, intracrystalline *vs.* intercrystalline)**

* Mostly <u>modular</u> (multidomains), ‘chimeric proteins’ ➔ Multifunctional

* Functions deduced from sequence analysis and comparison with known proteins of clearly identified functions: domain sequences similar to those found in vertebrates (in particular in ECM)

* Functional assays / gene knock-down : on very few proteins



Skeletal proteins

Domains of clearly identified function

- Enzymatic
- Sugar-binding
- Protease inhibitor
- Ca-binding
- Immunity
- ECM-like

Domains of unknown / putative function (no homologues)

- All LCD or RLCDs
- Other non-LCDs

A summary…

FEATURES

- Skeletal matrix proteins => Diversity => Plasticity
- Modular architecture of sequences
- Abundance of LCDs and RLCDs

MACROEVOLUTIONARY TRENDS

- Skeletal matrix proteins = complex evolutionary histories
  - Early recruitment (Cambrian or before) + 2^{ary} evolutions in the different lineages
  - Recent recruitment (lineage-dependent) and / or fast evolution: RLCDs

  - No phylogenetic signal (at high taxonomic level) !!!