

ΕΙΔΙΚΑ ΜΑΘΗΜΑΤΑ ΠΕΡΙΒΑΛΛΟΝΤΙΚΗΣ ΜΙΚΡΟΒΙΟΛΟΓΙΑΣ

Σύγχρονες μέθοδοι ανάλυσης μικροβιωμάτων Μικροβιακή ποικιλότητα á la NGS

Επιμέλεια

Βασιλειάδης Σωτήριος

(Ομάδα Βιοτεχνολογίας Φυτών και Περιβάλλοντος)

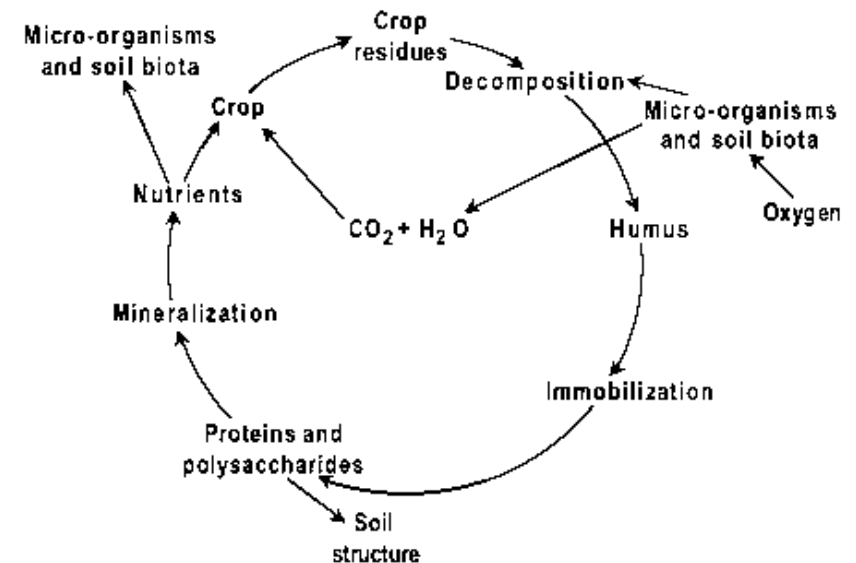
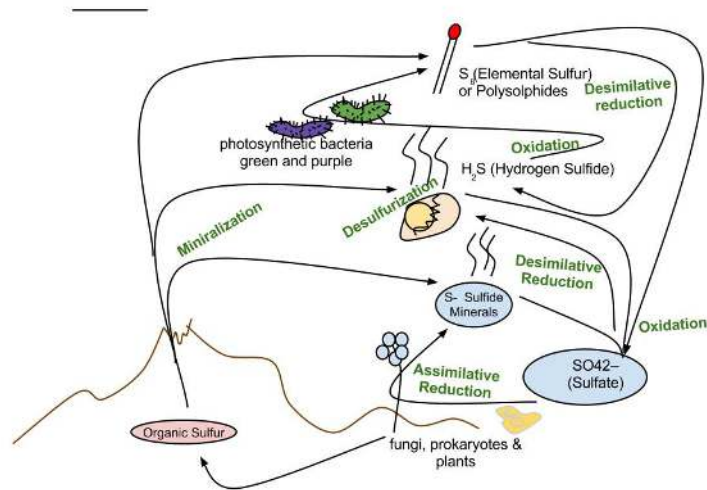
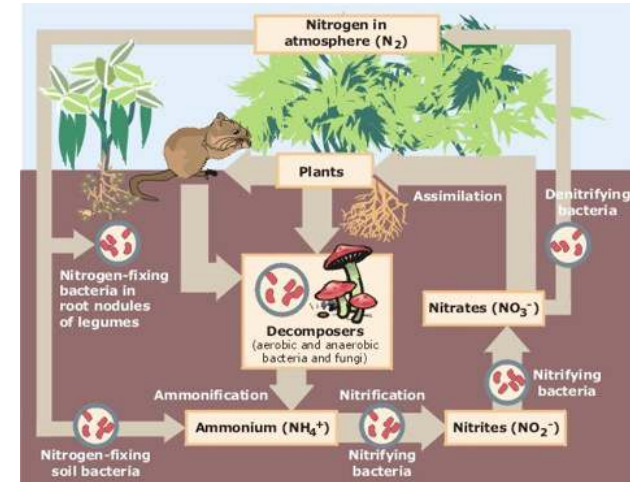


ΤΜΗΜΑ

Βιοχημείας &
Βιοτεχνολογίας
ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

Μικροβιακές παροχές οικοσυστημάτων

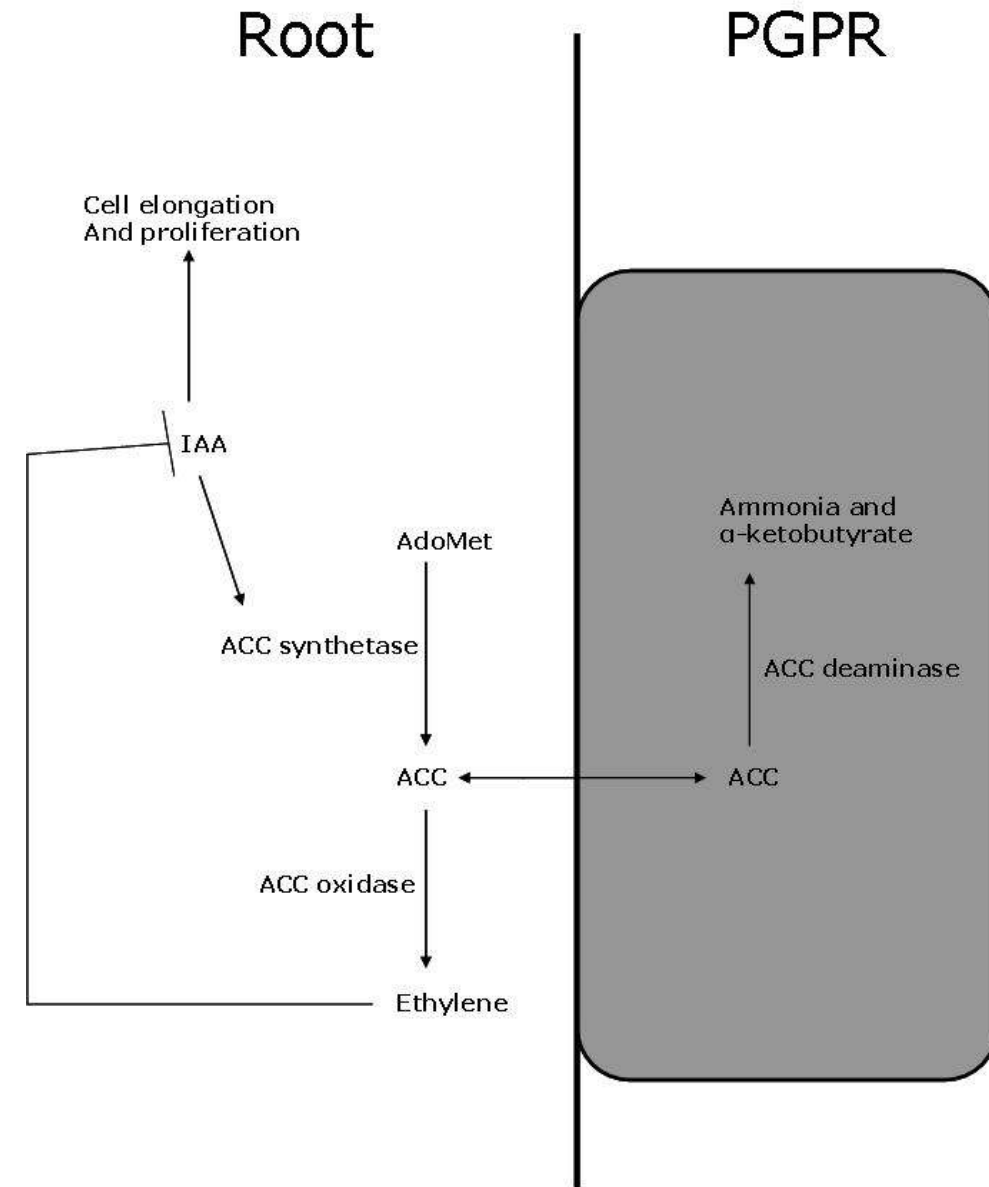
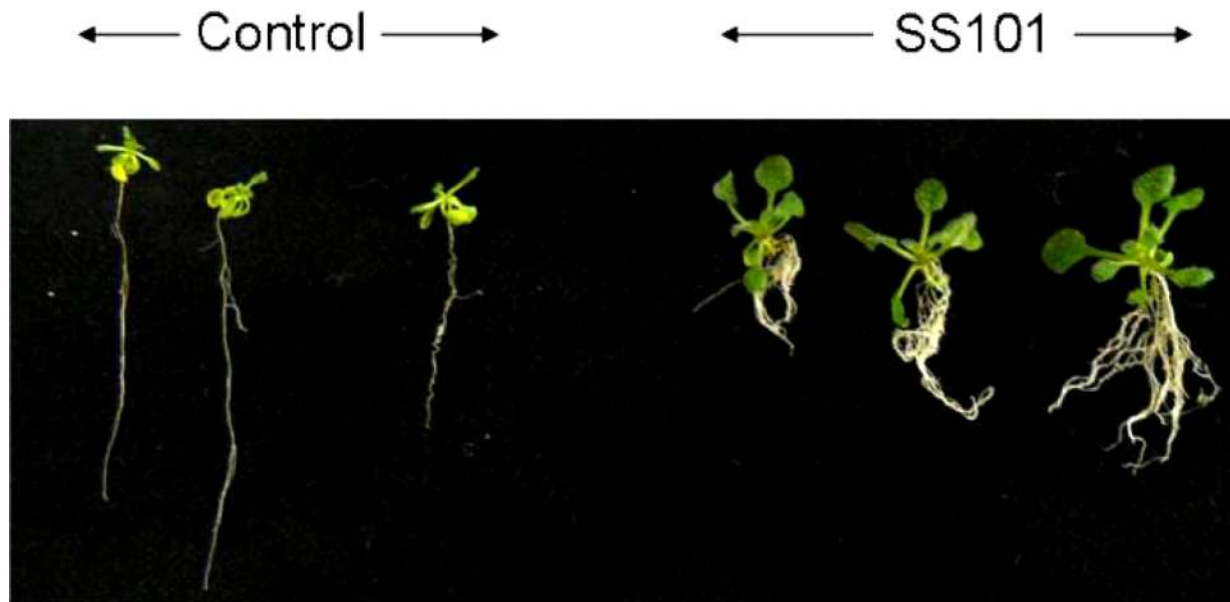
- Κύκλοι θρεπτικών
- Επίδραση σε αύξηση/ανάπτυξη φυτών ζώων
- Απορρύπανση στο πεδίο
- Βιο-έλεγχος παθογόνων, πηγή αντιβιοτικών, ενζύμων
- ...





Μικροβιακές παροχές οικοσυστημάτων

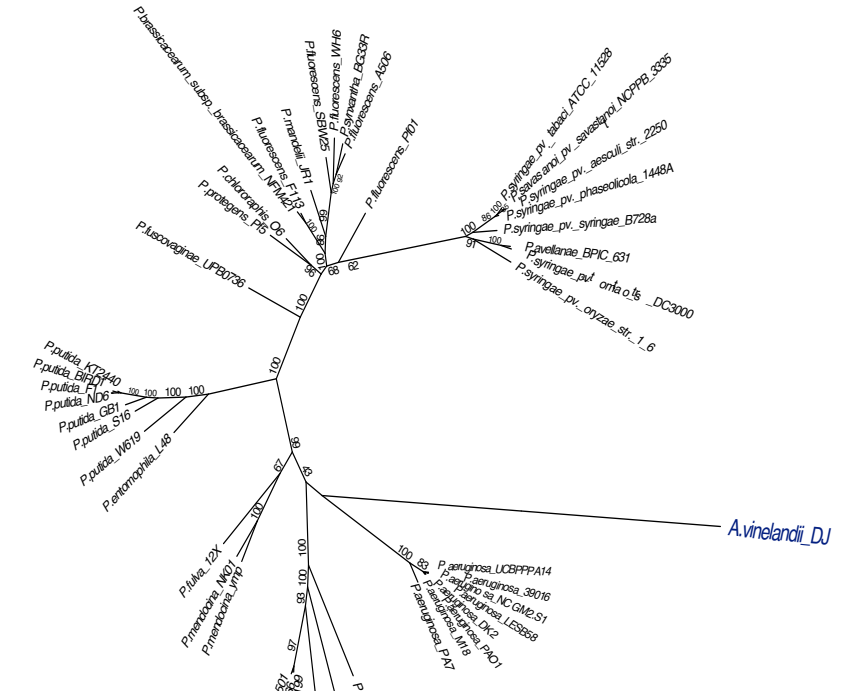
- Κύκλοι θρεπτικών
- Επίδραση σε αύξηση/ανάπτυξη φυτών ζώων
- Απορρύπανση στο πεδίο
- Βιο-έλεγχος παθογόνων, πηγή αντιβιοτικών, ενζύμων
- ...



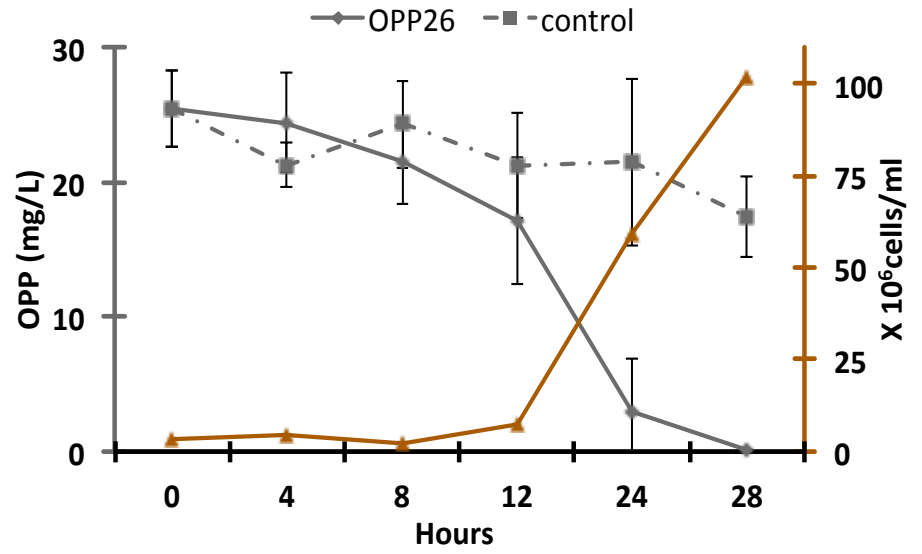
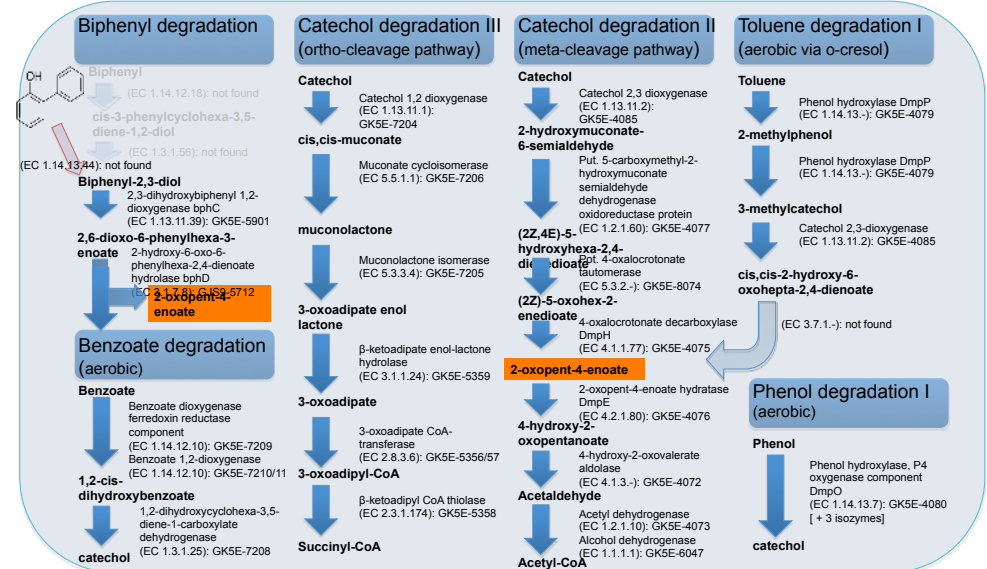


Μικροβιακές παροχές οικοσυστημάτων

- Κύκλοι θρεπτικών
- Επίδραση σε αύξηση/ανάπτυξη φυτών ζώων
- Απορρύπανση στο πεδίο
- Βιο-έλεγχος παθογόνων, πηγή αντιβιοτικών, ενζύμων



• ...





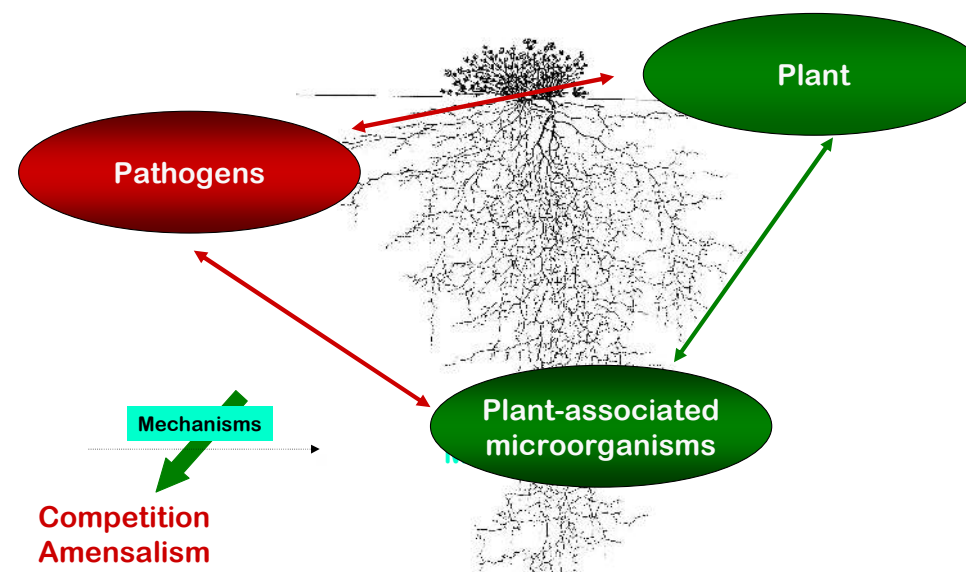
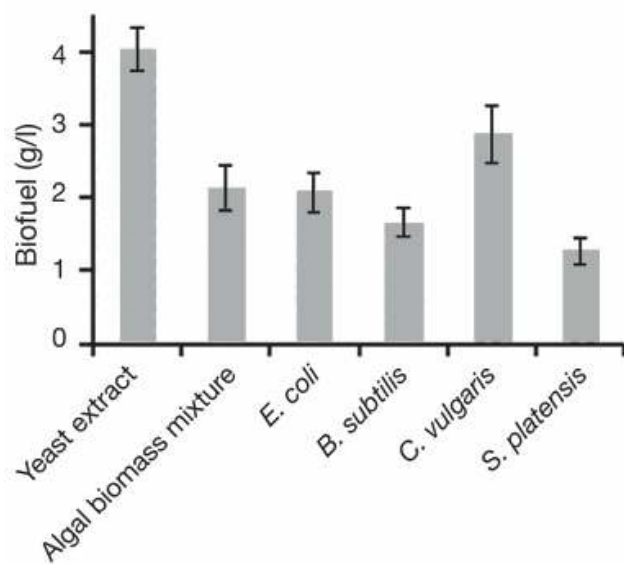
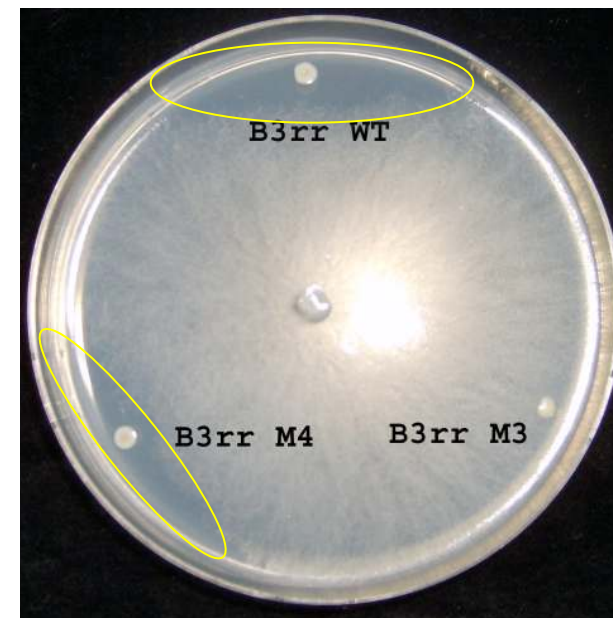
ΤΜΗΜΑ

Βιοχημείας &
Βιοτεχνολογίας

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

Μικροβιακές παροχές οικοσυστημάτων

- Κύκλοι θρεπτικών
- Επίδραση σε αύξηση/ανάπτυξη φυτών ζώων
- Απορρύπανση στο πεδίο
- Βιο-έλεγχος παθογόνων, πηγή αντιβιοτικών, ενζύμων
- ...





ΤΜΗΜΑ

Βιοχημείας &
Βιοτεχνολογίας

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

Μικροοργανισμοί σε αριθμούς

- 10^9 κύτταρα και $1-5 \cdot 10^4$ στελέχη ανά γραμμάριο εύκρατου επιφανειακού εδάφους
- 10^4 στελέχη στο ανθρώπινο σώμα (τα προκαρυωτικά κύτταρα αποτελούν το 70 % του αριθμού των κυττάρων που βρίσκονται στο ανθρώπινο σώμα).
- 100 εκατομμύρια περισσότεροι μικροοργανισμοί στους οκεανούς από τα αστέρια του γνωστού διαστήματος
- 400 γραμμάρια νευροτοξίνης του *Clostridium botulinum* μπορεί να εξαλείψει την ανθρωπότητα

EDITORIAL

Microbiology by numbers

The scale of life in the microbial world is such that amazing numbers become commonplace. These numbers can be sources of inspiration for those in the field and used to inspire awe in the next generation of microbiologists.

Nature E. 2011. Microbiology by numbers. *Nature* **9**:628-628.

Dance A. 2008. Soil ecology: What lies beneath. *Nature* **455**:724-725.
And citations therein...

What lies beneath

More creatures live in soil than any other environment on Earth. But what are they all doing there? **Amber Dance** reports on the world's widest biodiversity.

terial cells each day. Moving onto dry land, the number of microorganisms in a teaspoon of soil (1×10^9) is the same as the number of humans currently living in Africa. Even more amazingly, dental plaque is so dense

...The first DNA-based estimate of soil microbial biodiversity, published in 1990, counted about 4,000 different bacterial genomes per gram of soil. Since then, various studies and models have pushed the number up as high as 830,000 species per gram, down to 2,000, and back up again. Most recently, Triplett and his colleagues ran 139,000 individual sequences — more than other studies have used — and came up with an estimate of 10,000 to 50,000 species per gram of soil...



ΤΜΗΜΑ

Βιοχημείας &
Βιοτεχνολογίας

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

«Πάρτι» εδάφους και ριζόσφαιρας





ΤΜΗΜΑ

Βιοχημείας &
Βιοτεχνολογίας

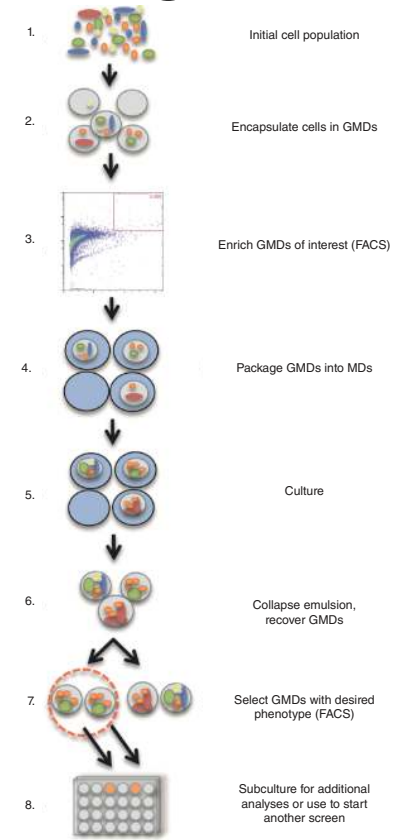
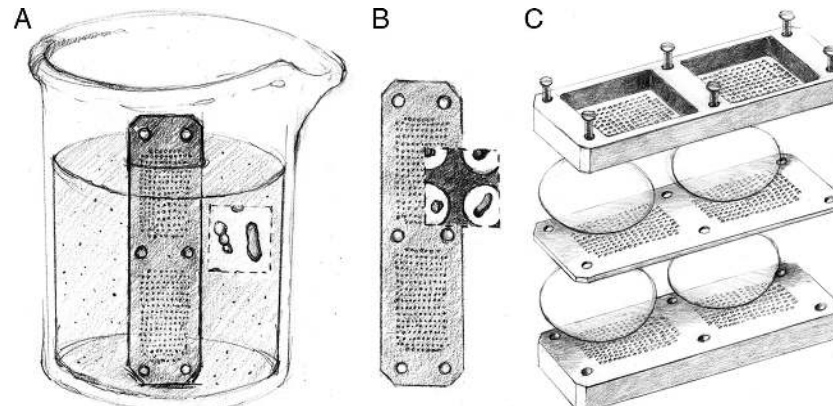
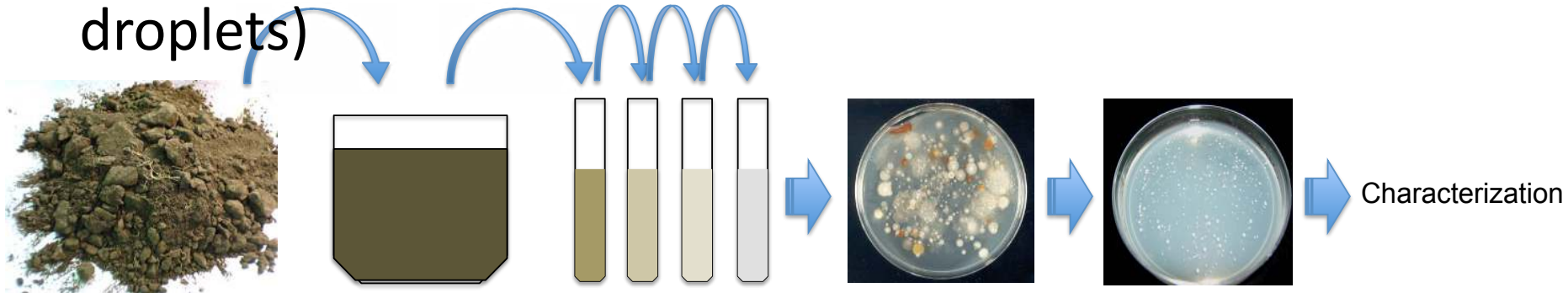
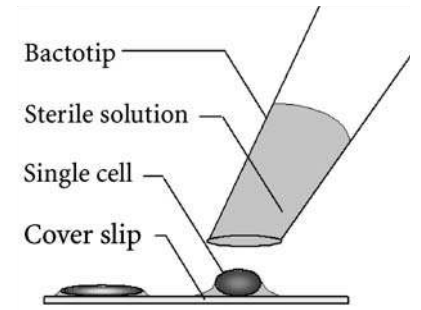
ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

Μέθοδοι ανάλυσης μικροβιωμάτων

- Καλλιεργητικές μέθοδοι
- Μέθοδοι ελεύθερες καλλιέργειας

Καλλιεργητικές μέθοδοι

- Κλασσικές
- Μεγάλης απόδοσης μέθοδοι (ρομποτικά colony pickers)
- Αντιμετώπισης ανταγωνισμού (i-Chip, micromanipulation, microbial gel droplets)



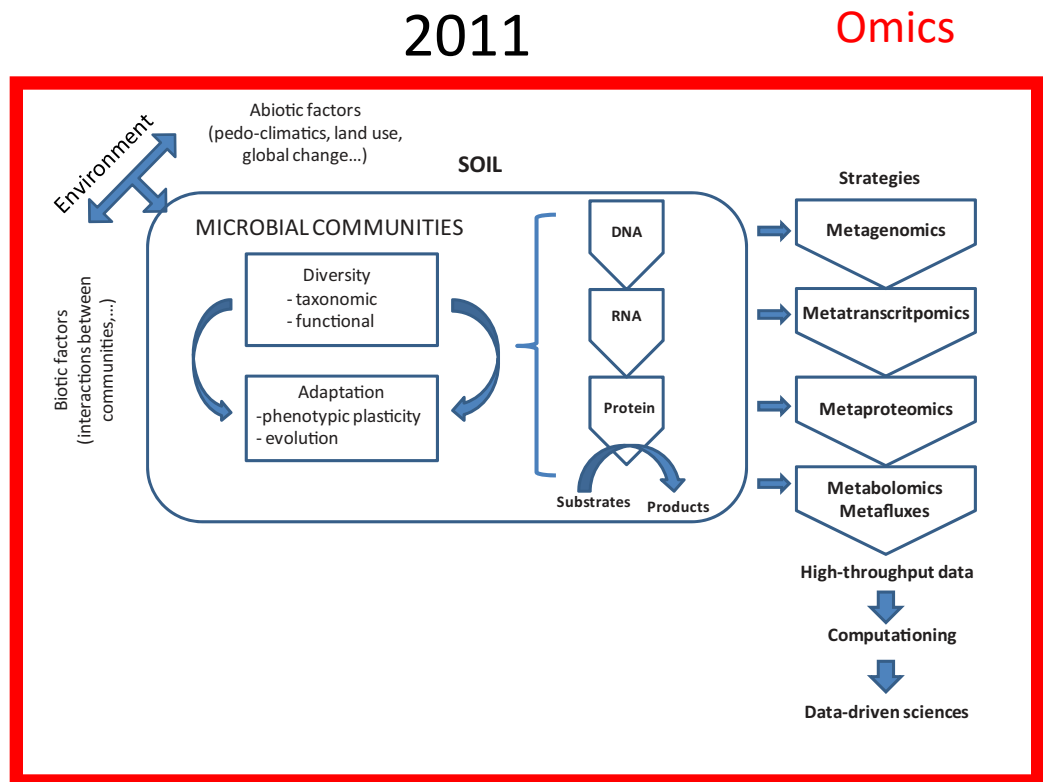
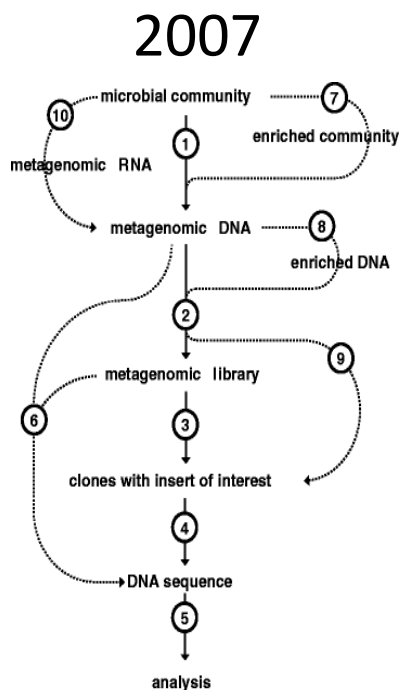
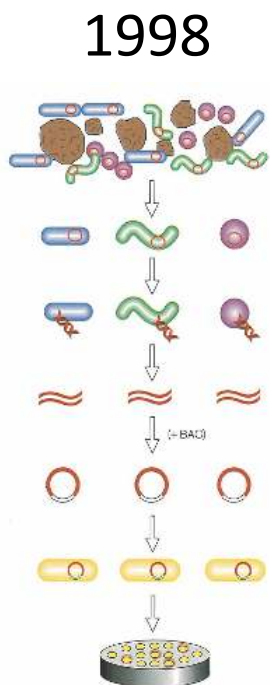


ΤΜΗΜΑ

Βιοχημείας &
Βιοτεχνολογίας

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

Μέθοδοι ελεύθερες καλλιέργειας










Μεταγονιδιωματική

“Η ανάλυση σε επίπεδο λειτουργίας και σύνθεσης αλληλουχιών DNA του συνόλου των γονιδιωμάτων περιβαλλοντικών δειγμάτων”

Μέθοδοι ελεύθερες καλλιέργειας (ποικιλότητα)

- Υβριδισμός (ιχνηλάτες) πχ:
 - Μικροσκοπία (φθορίζοντος *in situ* υβριδισμός - FISH)
 - Κυτταρομετρητές ροής
 - Μικρο-συστοιχίες (phylochip)
- Αλυσιδωτή αντίδραση πολυμεράσης (PCR) πχ:
 - DGGE
 - T-RFLP, ARISA
 - Μέθοδοι Αλληλούχισης νέας γενιάς

Category	Hybridization based			PCR based			
Method	FISH, CARD FISH		Microarrays	DGGE	T-RFLP, ARISA	qPCR	HTS
	microscopy	Flow cytometry					
Traits							
Analyzed samples per day or run	Tens per day	Tens per day	One per run	Tens per day	Hundreds per day	Hundreds per day	Hundreds per run
Signatures simultaneously screened per sample	Less than 10	Less than 10	Thousands, to hundreds of thousands	Tens	hundreds	One	Hundreds, to hundreds of thousands, to millions
Qualitative / Quantitative	Quantitative	Quantitative	Quantitative **	Semi-quantitative	Quantitative *	Quantitative	Quantitative *
Automated	No	Yes	Yes	No	Yes	Yes	Yes
Confirmation tests or suggested experimental validation	Double probing	Double probing	qPCR	Cloning and sequencing	Cloning, clone to polymorphisms match, clone sequencing	No	No

*: careful preparation (e.g. low number of PCR cycles) is necessary for reducing intensity of the PCR plateau effect on quantification abilities

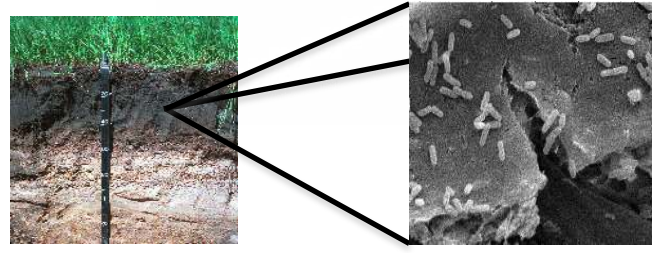
** : the PCR plateau effect introduced bias is applicable in case a single marker gene like the SSU is screened through multiple taxa after PCR

Σε περιπτώσεις που ενδιαφερόμαστε για φυλογενετικά μοτίβα πολυποίκιλων μικροβιακών ομάδων περιβαλλοντικών δειγμάτων συχνά καταφεύγουμε σε

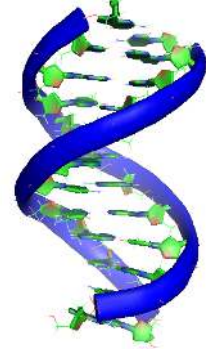
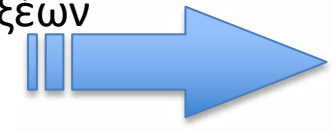
Ανάλυσεις προϊόντων PCR με μεθόδους αλληλούχισης νέας (NGS)



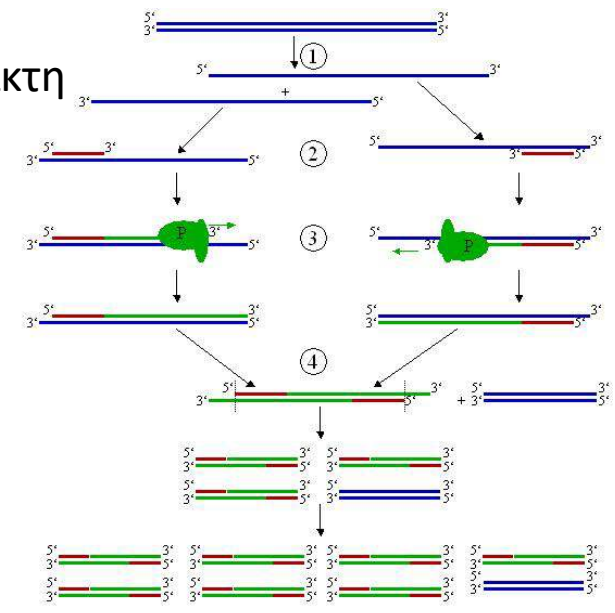
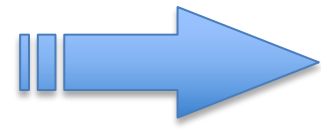
Μικροβιακή ποικιλότητα με NGS με μια ματιά



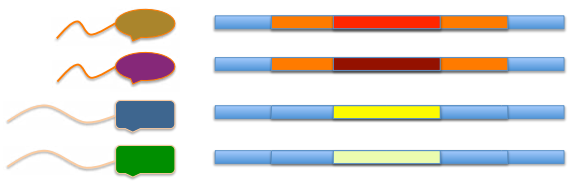
1. Δειγματοληψία και εκχύλιση νουκλεϊκών οξέων



2. Επιλογή και PCR ενίσχυση φυλογενετικού δείκτη



Ταξινομικές μονάδες



4. Ανάλυση δεδομένων



3. Αλληλούχιση

Ανάλυσεις προϊόντων PCR με μεθόδους Αλληλούχισης νέας (NGS)

1. Δειγματοληψία και εκχύλιση ν. ο.



ΤΜΗΜΑ

Βιοχημείας &
Βιοτεχνολογίας

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

1. Δειγματοληψία και εκχύλιση ν. ο.

- Δειγματοληψία
 - Αντιπροσωπευτικό δείγμα
 - Αντιπροσωπευτικός αριθμός επαναλήψεων
 - Μάρτυρες!!!
 - Μέτρηση κατάλληλων παραμέτρων.
 - ...
- Εκχύλιση νουκλεϊκών οξέων... βήματα:
 - Διάρρηξη κυτταρικών μεμβρανών/τοιχωμάτων
 - Αδρανοποίηση αποδιάταξη νουκλεασών
 - Καθαρισμός από πρωτεΐνες και παρεμποδιστές PCR

Ανάλυσεις προϊόντων PCR με μεθόδους Αλληλούχισης νέας (NGS)

2α. Επιλογή φυλογενετικού δείκτη



ΤΜΗΜΑ

Βιοχημείας &
Βιοτεχνολογίας

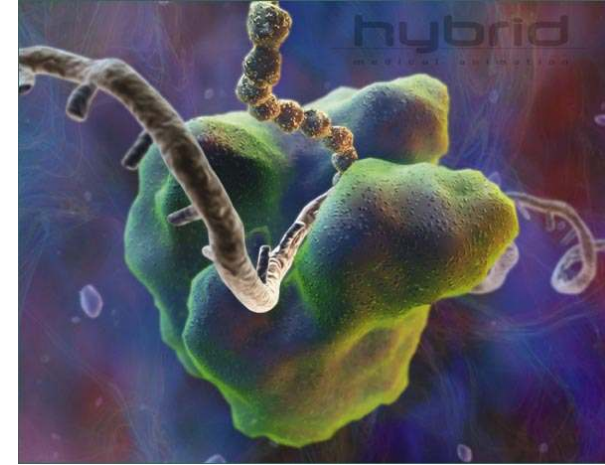
ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

2α. Επιλογή φυλογενετικού δείκτη

- Να υπάρχει στους μικροοργανισμούς ενδιαφέροντος:
 - 16S rRNA γονίδιο: βακτήρια, αρχαία
 - ITS: μύκητες
 - 18S rRNA γονίδιο: πρώτιστα, μυκόριζες, ευκαρυώτες
 - amoA: νιτροδοποιητές
 - ...
- Να υπάρχουν συντηρημένες αλληλουχίες DNA κατάλληλες για σχεδιασμό εκκινητών
- Να παράγονται προϊόντα PCR συμβατά με την τεχνολογία αλληλούχισης
- Να υπάρχει αρκετή παραλλακτικότητα στις αλληλουχίες στόχους -> κατάταξη σε χαμηλές ταξινομικές βαθμίδες

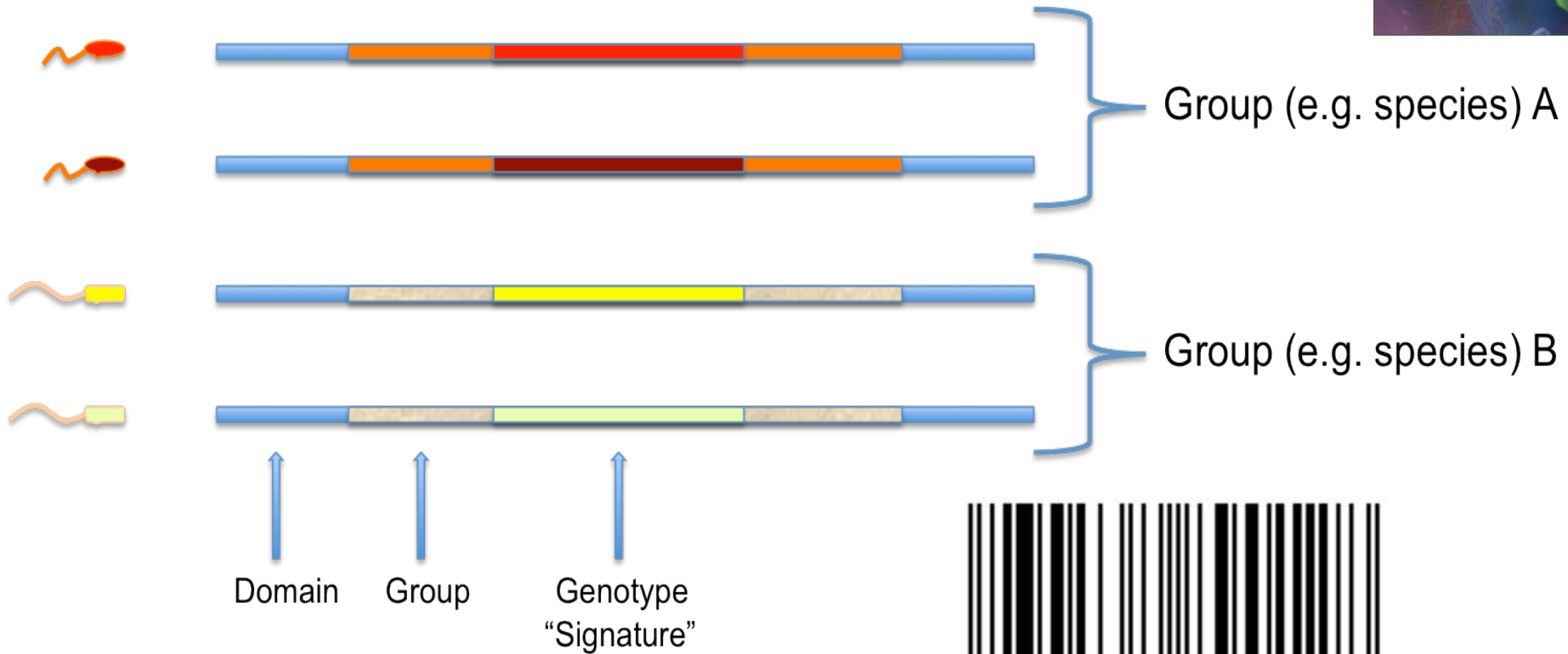


2α. Π.χ. το 16S rRNA γονίδιο

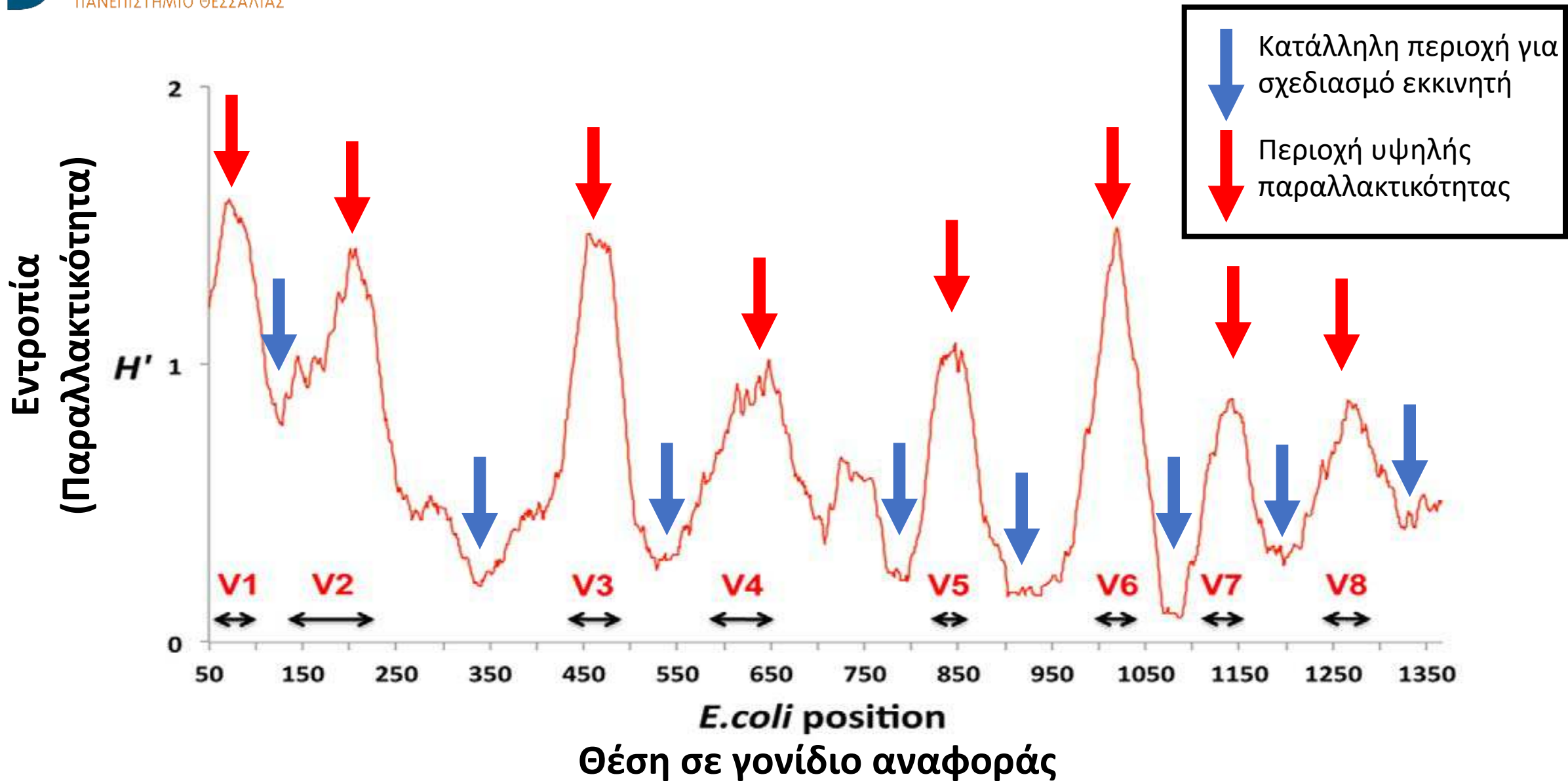


Microorganisms

Bacterial 16S rRNA strand fragment

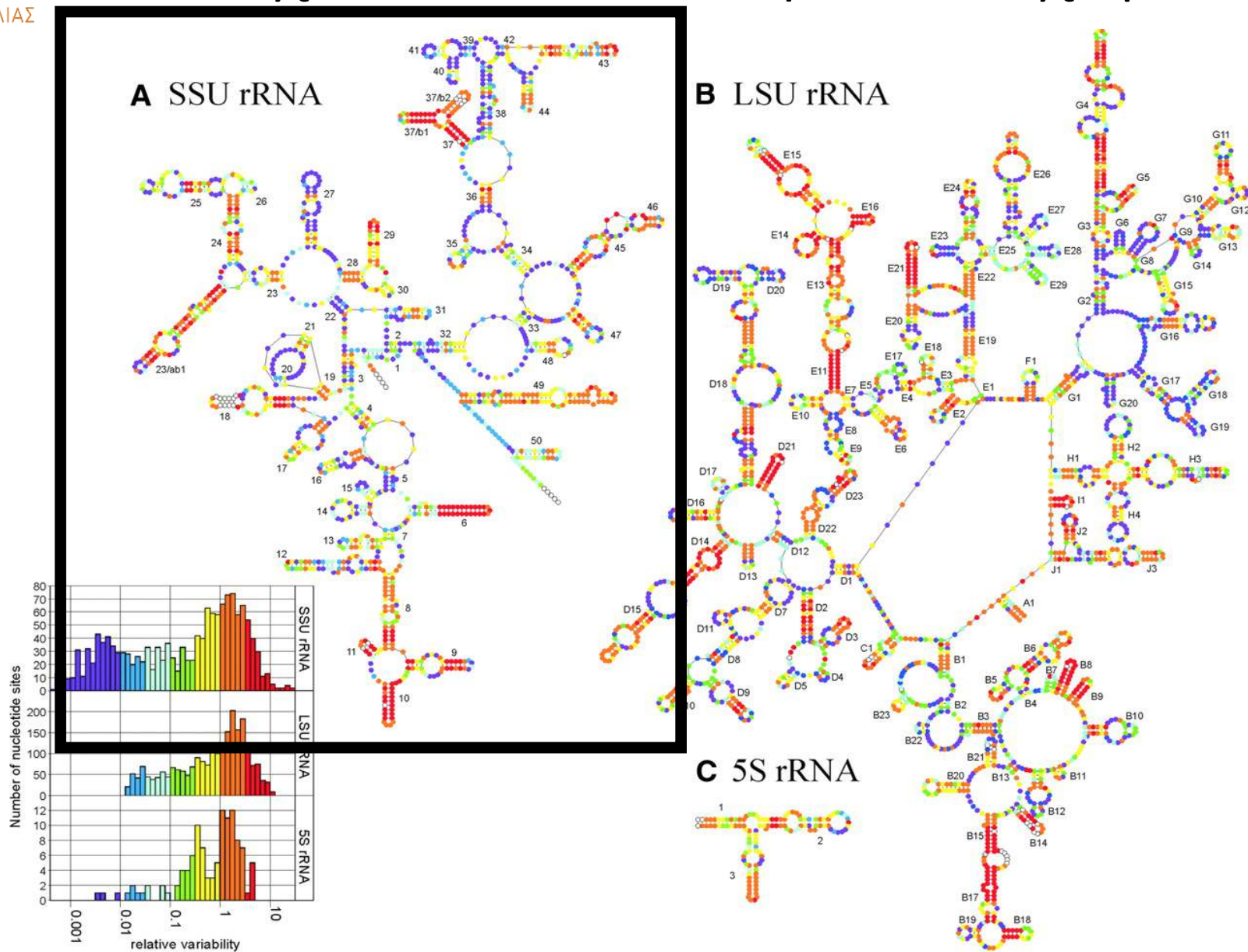


2α. Π.χ. το 16S rRNA γονίδιο





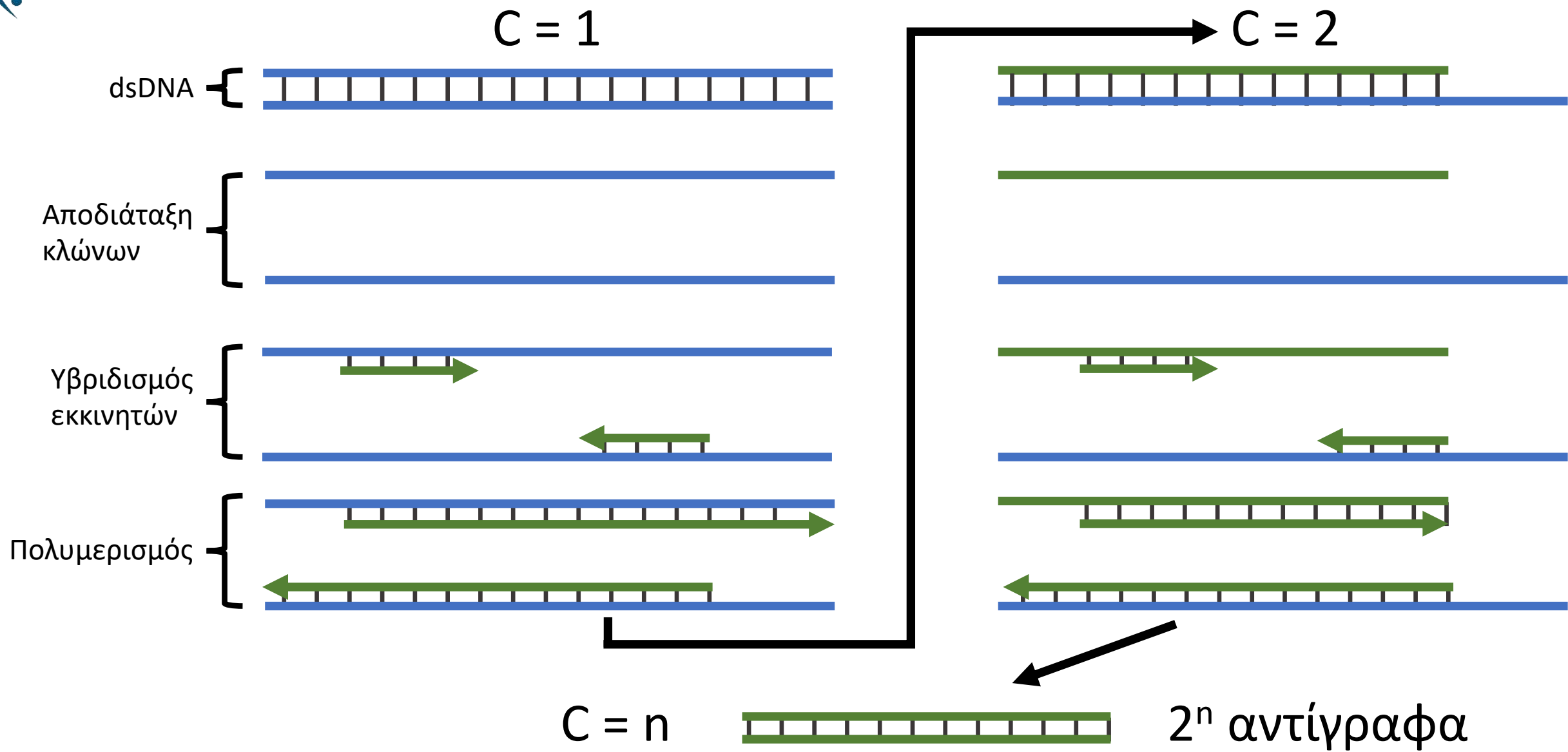
2α. Π.χ. το 16S rRNA γονίδιο χάρτες εντροπίας



Ανάλυσεις προϊόντων PCR με μεθόδους Αλληλούχισης νέας (NGS)

2β. Αλυσιδωτή αντίδραση πολυμεράσης (PCR)

2β. PCR (βασικές έννοιες)





ΤΜΗΜΑ

Βιοχημείας &
Βιοτεχνολογίας

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

2β. λάθη PCR & πρόληψη/αντιμετώπιση

*Γενικότερα το βάθος της Αλληλούχισης αντισταθμίζει τα κόστη (όσο περισσότερα τα αναγνώσματα τόσο περισσότερα λάθη μπορούμε να έχουμε)

- Λάθη -> πρόληψη/αντιμετώπιση :

- ❖ Λανθασμένος υβριδισμός εκκινητών -> θερμοκρασία/συνθήκες υβριδισμού & *

- ❖ Χιμαιρικά προϊόντα ενίσχυσης -> υψηλής πιστότητας πολυμεράσες, ελαχιστοποίηση κύκλων PCR & *

- ❖ Ενίσχυση ομόλογων αλληλουχιών μη-στόχων ως προς την ανάλυση (π.χ. στην περίπτωση του 16S rRNA γονιδίου, το μιτοχονδριακό ομόλογο ή το ομόλογο των χλωροπλαστών) -> αντιπροσωπευτικές μέθοδοι εκχύλισης, γνώση περιβάλλοντος & *



ΤΜΗΜΑ

Βιοχημείας &
Βιοτεχνολογίας

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

2β. λάθη PCR & πρόληψη/αντιμετώπιση

*Γενικότερα το βάθος της Αλληλούχισης αντισταθμίζει τα κόστη (όσο περισσότερα τα αναγνώσματα τόσο περισσότερα λάθη μπορούμε να έχουμε)

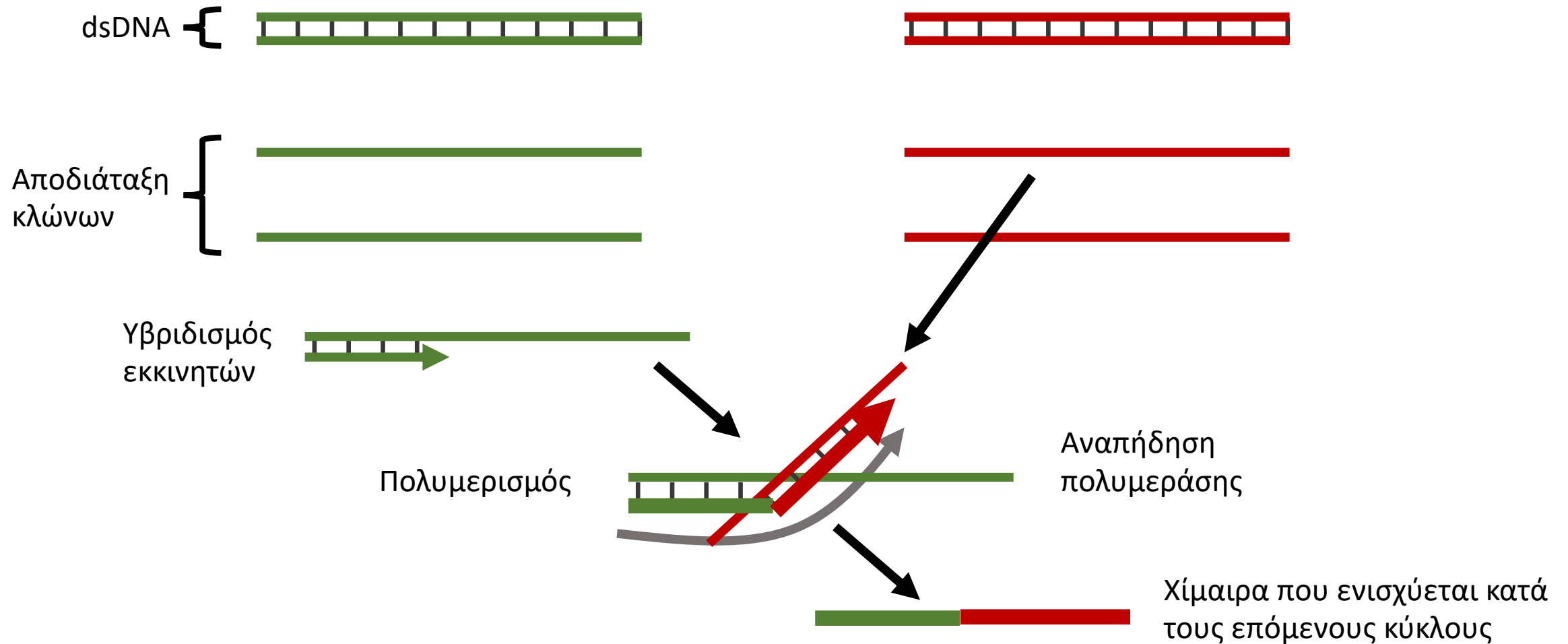
- Λάθη -> πρόληψη/αντιμετώπιση :

- ❖ Λανθασμένος υβριδισμός εκκινητών -> θερμοκρασία/συνθήκες υβριδισμού & *

- ❖ Χιμαιρικά προϊόντα ενίσχυσης -> υψηλής πιστότητας πολυμεράσες, ελαχιστοποίηση κύκλων PCR & *

- ❖ Ενίσχυση ομόλογων αλληλουχιών μη-στόχων ως προς την ανάλυση (π.χ. στην περίπτωση του 16S rRNA γονιδίου, το μιτοχονδριακό ομόλογο ή το ομόλογο των χλωροπλαστών) -> αντιπροσωπευτικές μέθοδοι εκχύλισης, γνώση περιβάλλοντος & *

2β. λάθη PCR (Χίμαιρες)





ΤΜΗΜΑ

Βιοχημείας &
Βιοτεχνολογίας

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

2β. λάθη PCR & πρόληψη/αντιμετώπιση

*Γενικότερα το βάθος της Αλληλούχισης αντισταθμίζει τα κόστη (όσο περισσότερα τα αναγνώσματα τόσο περισσότερα λάθη μπορούμε να έχουμε)

- Λάθη -> πρόληψη/αντιμετώπιση :

- ❖ Λανθασμένος υβριδισμός εκκινητών -> θερμοκρασία/συνθήκες υβριδισμού & *

- ❖ Χιμαιρικά προϊόντα ενίσχυσης -> υψηλής πιστότητας πολυμεράσες, ελαχιστοποίηση κύκλων PCR & *

- ❖ Ενίσχυση ομόλογων αλληλουχιών μη-στόχων ως προς την ανάλυση (π.χ. στην περίπτωση του 16S rRNA γονιδίου, το μιτοχονδριακό ομόλογο ή το ομόλογο των χλωροπλαστών) -> αντιπροσωπευτικές μέθοδοι εκχύλισης, γνώση περιβάλλοντος & *



ΤΜΗΜΑ

Βιοχημείας &
Βιοτεχνολογίας

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

2β. λάθη PCR (Χίμαιρες: αντιμετώπιση)

Αντιμετώπιση χιμαιρών κατά την ανάλυση δεδομένων:

- Αντιπαραβολή του κάθε μισού του αναγνώσματος με βάση δεδομένων:
 - εμπιστοσύνης
 - εκ νέου κατασκευασμένης από τα δεδομένα (ο κανόνας του πιο άφθονου)
- Αν η αλληλουχία περιέχει δύο μισά που απαντώνται σε διαφορετικές ταξινομικές μονάδες κατατάσσεται στις χιμαιρικές



ΤΜΗΜΑ

Βιοχημείας &
Βιοτεχνολογίας

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

2β. λάθη PCR & πρόληψη/αντιμετώπιση

*Γενικότερα το βάθος της Αλληλούχισης αντισταθμίζει τα κόστη (όσο περισσότερα τα αναγνώσματα τόσο περισσότερα λάθη μπορούμε να έχουμε)

• Λάθη -> πρόληψη/αντιμετώπιση :

- ❖ Λανθασμένος υβριδισμός εκκινητών -> θερμοκρασία/συνθήκες υβριδισμού & *
- ❖ Χιμαιρικά προϊόντα ενίσχυσης -> υψηλής πιστότητας πολυμεράσες, ελαχιστοποίηση κύκλων PCR & *
- ❖ Ενίσχυση ομόλογων αλληλουχιών μη-στόχων ως προς την ανάλυση (π.χ. στην περίπτωση του 16S rRNA γονιδίου, το μιτοχονδριακό ομόλογο ή το ομόλογο των χλωροπλαστών) -> αντιπροσωπευτικές μέθοδοι εκχύλισης, γνώση περιβάλλοντος & *

Ανάλυσεις προϊόντων PCR με μεθόδους αλληλούχισης νέας (NGS)

3. Αλληλούχιση

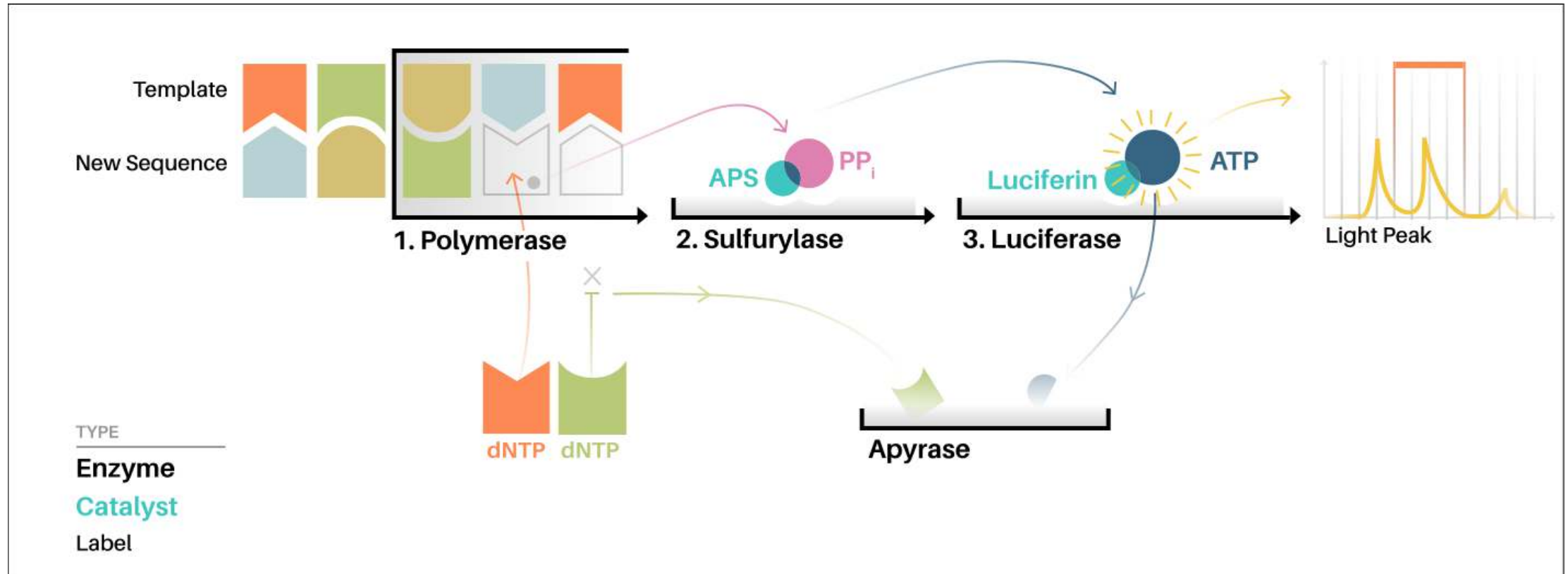


ΤΜΗΜΑ

Βιοχημείας &
Βιοτεχνολογίας

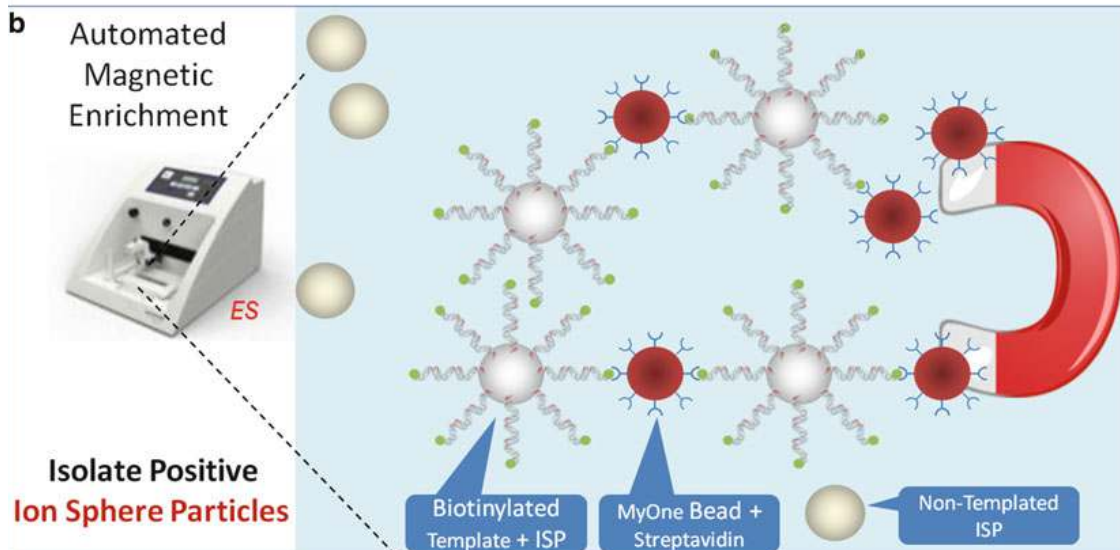
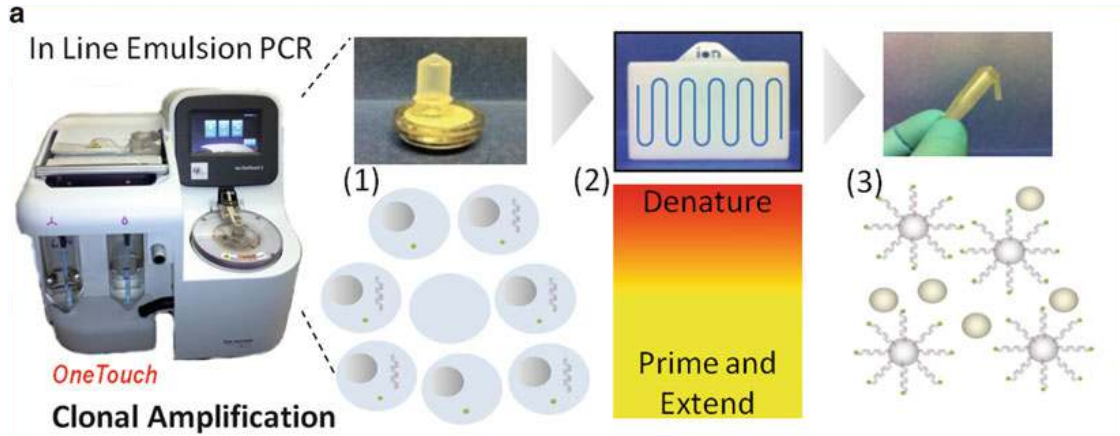
ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

3. 2^{ης} γενιάς Pyrosequencing



3. 2^ης γενιάς Ion Torrent/Proton

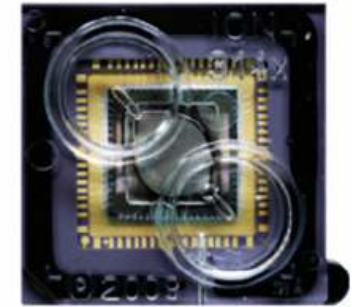
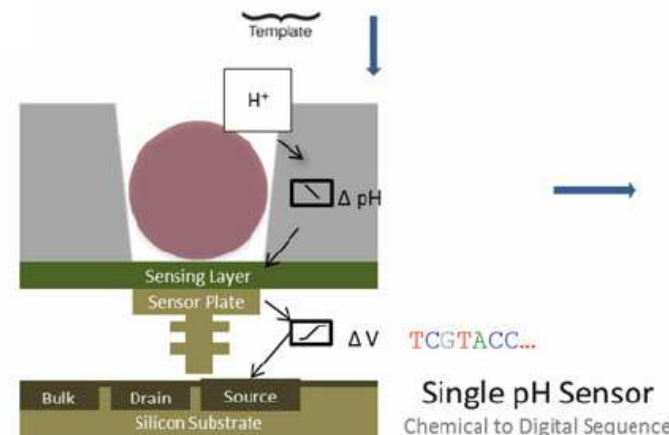
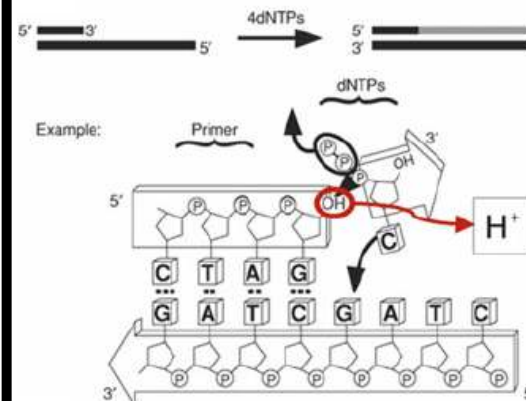
Sample prep



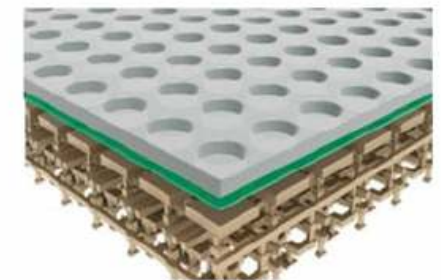
Sequencing

Principle and Elements of Semiconductor Sequencing

Simple Natural Chemistry of Sequencing-by-Synthesis with H⁺ release detection



Sequencing Chip
Semiconductor Packaging

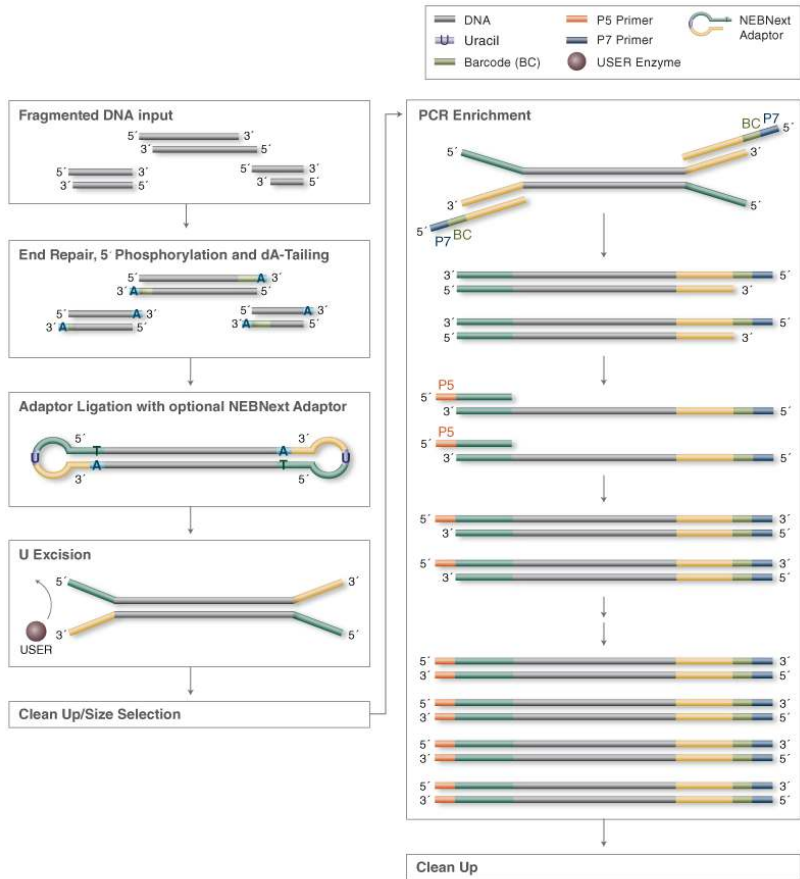


Millions of pH Sensors
Semiconductor Design

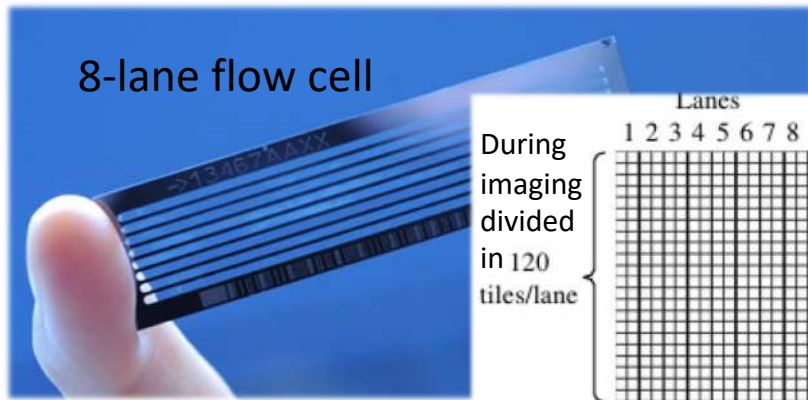
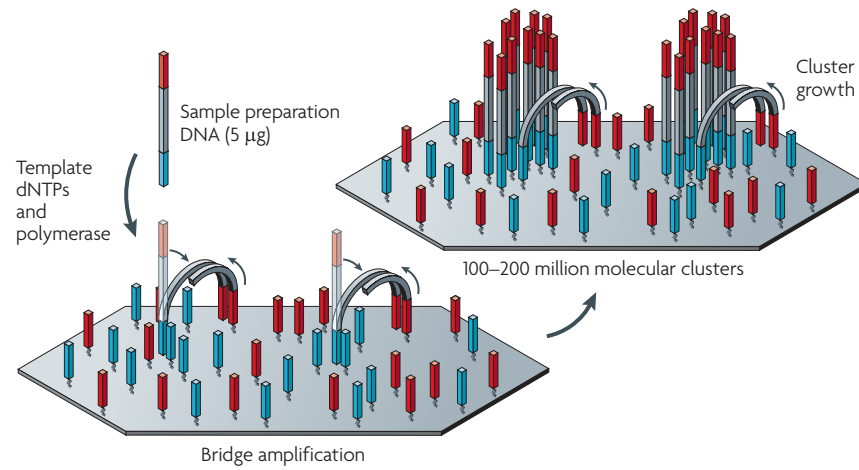


3. 2^ης γενιάς Illumina

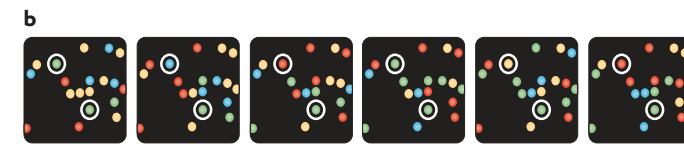
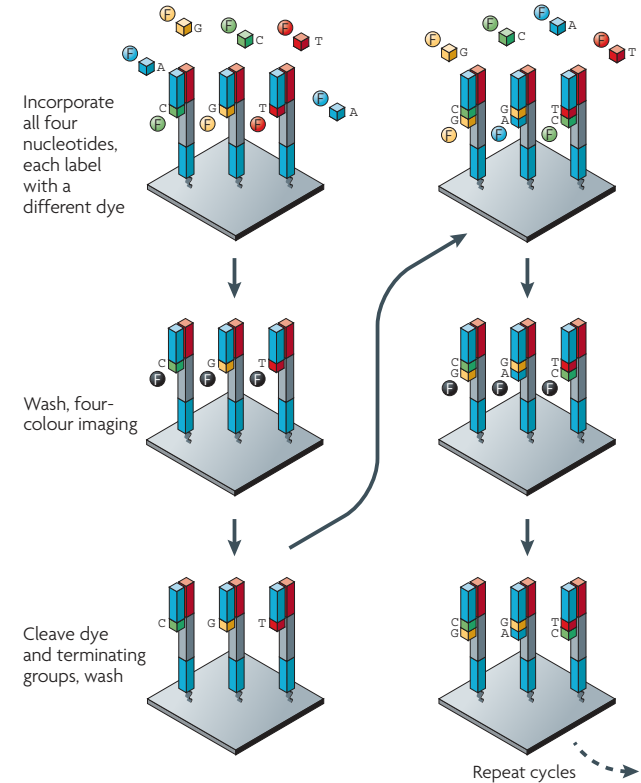
Library prep



Flow cell attachment and bridge amplification



Sequencing data generation



Top: CATCGT
Bottom: CCCCC

<https://www.youtube.com/watch?v=fCd6B5HRaZ8>

<https://www.neb.com/applications/library-preparation-for-next-generation-sequencing/illumina-library-preparation>

Kawashima, E., *et al.* (1998). Method of nucleic acid amplification, Google Patents

Metzker, M.L. (2009). Sequencing technologies — the next generation. *Nat Rev Genet* 11, 31 (modified)



3. Πιθανότητα λάθους (Phred Q values)

$$Q = -10 \times \log_{10}(P_{err}) \quad \Leftrightarrow \quad P_{err} = 10^{Q/-10}$$

Representation of quality scores

Table 1 ASCII Characters Encoding Q-scores 0-40

Symbol	ASCII Code	Q-Score	Symbol	ASCII Code	Q-Score	Symbol	ASCII Code	Q-Score
!	33	0	/	47	14	=	61	28
"	34	1	0	48	15	>	62	29
#	35	2	1	49	16	?	63	30
\$	36	3	2	50	17	@	64	31
%	37	4	3	51	18	A	65	32
&	38	5	4	52	19	B	66	33
'	39	6	5	53	20	C	67	34
(40	7	6	54	21	D	68	35
)	41	8	7	55	22	E	69	36
*	42	9	8	56	23	F	70	37
+	43	10	9	57	24	G	71	38
,	44	11	:	58	25	H	72	39
-	45	12	;	59	26	I	73	40
.	46	13	<	60	27			

Phred quality scores are logarithmically linked to error probabilities

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%
60	1 in 1,000,000	99.9999%

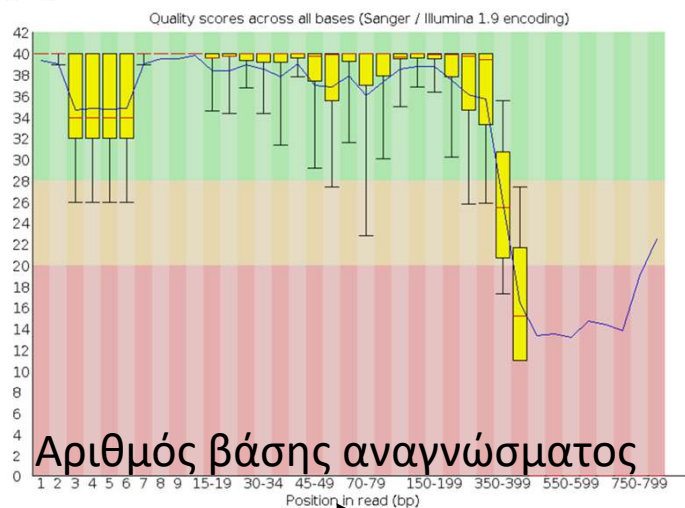


3. Πιθανότητα λάθους ανά θέση αναγνώσματος

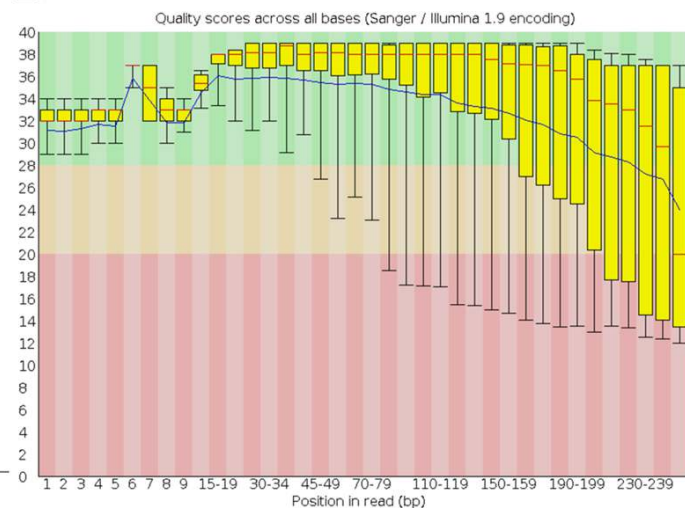
Μέση Phred Q

Αριθμός αναγνωσμάτων

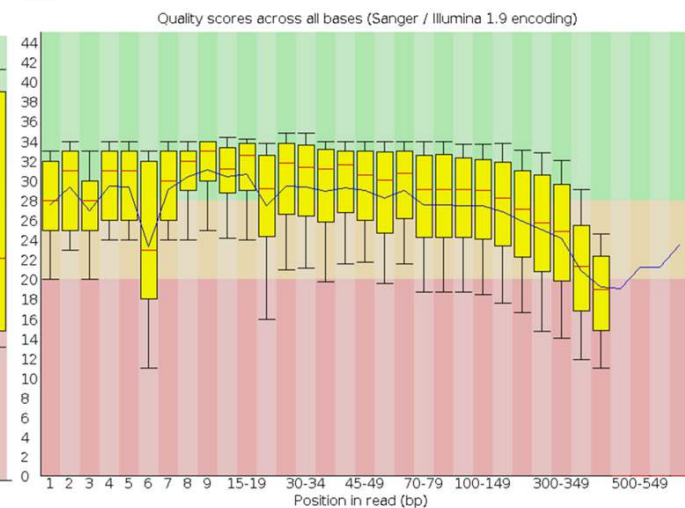
A Pyrosequencing



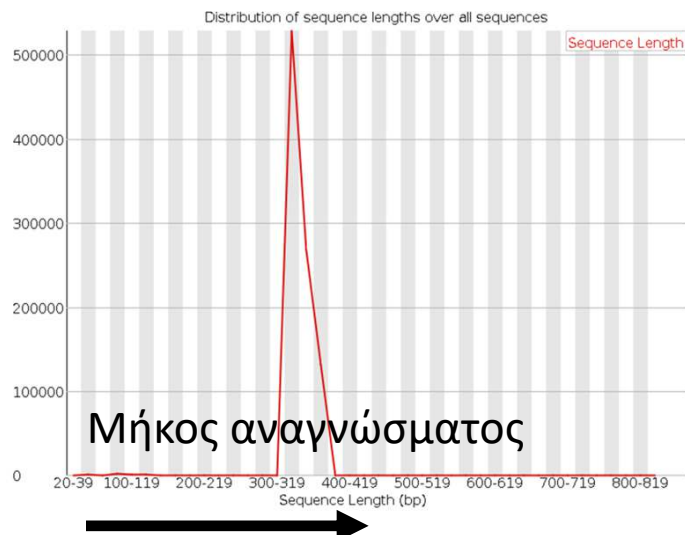
B Illumina



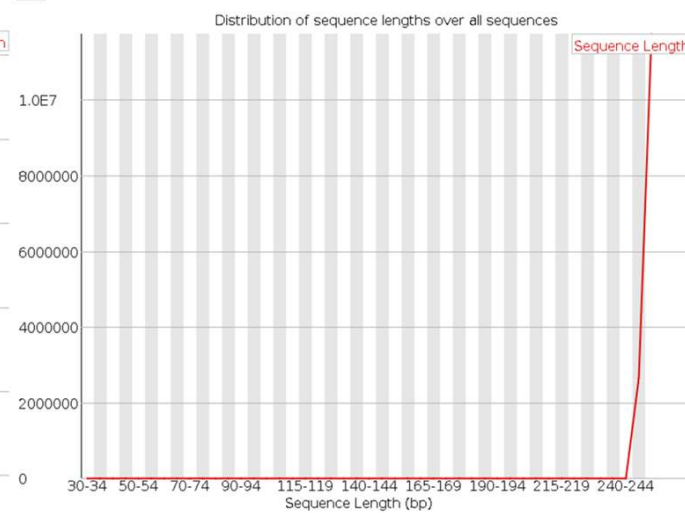
C Ion Torrent/Proton



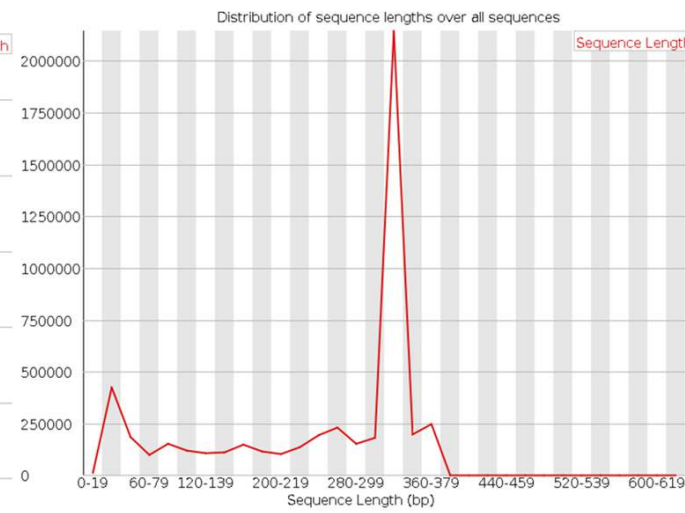
D



E



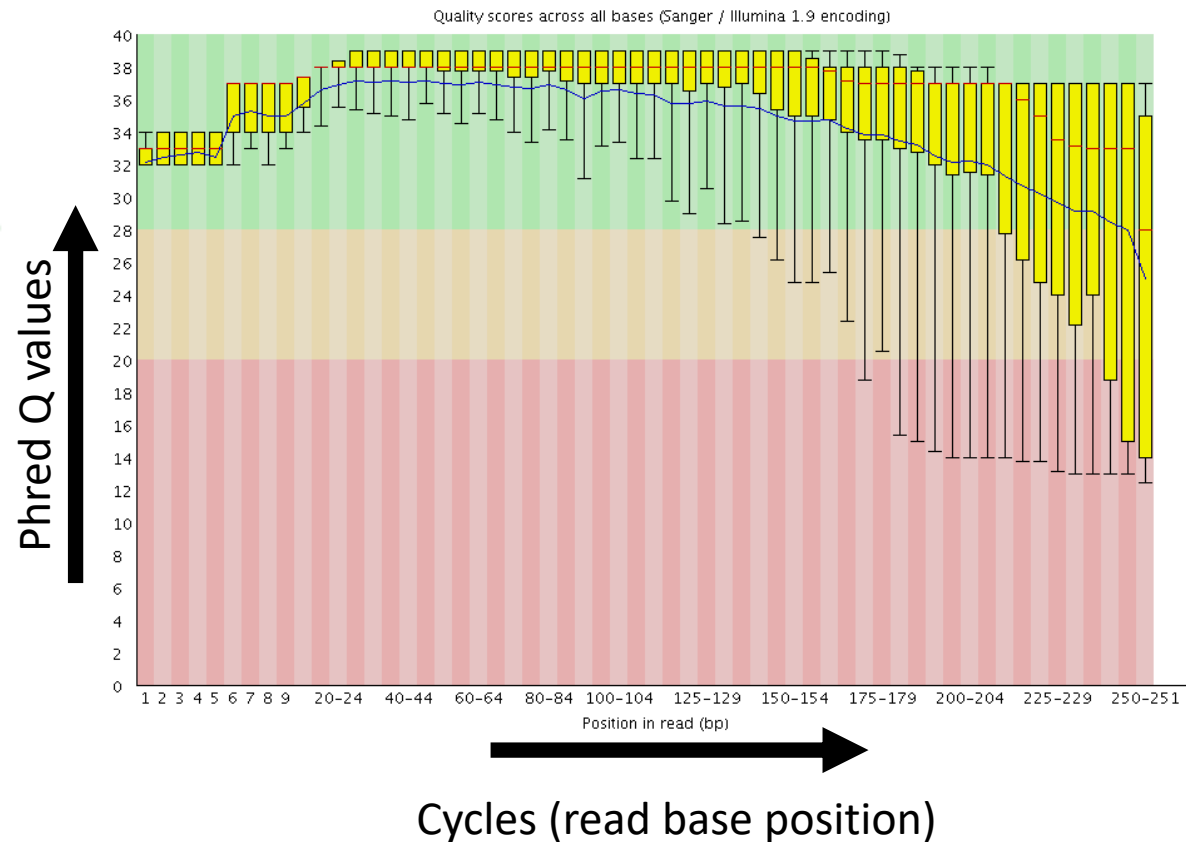
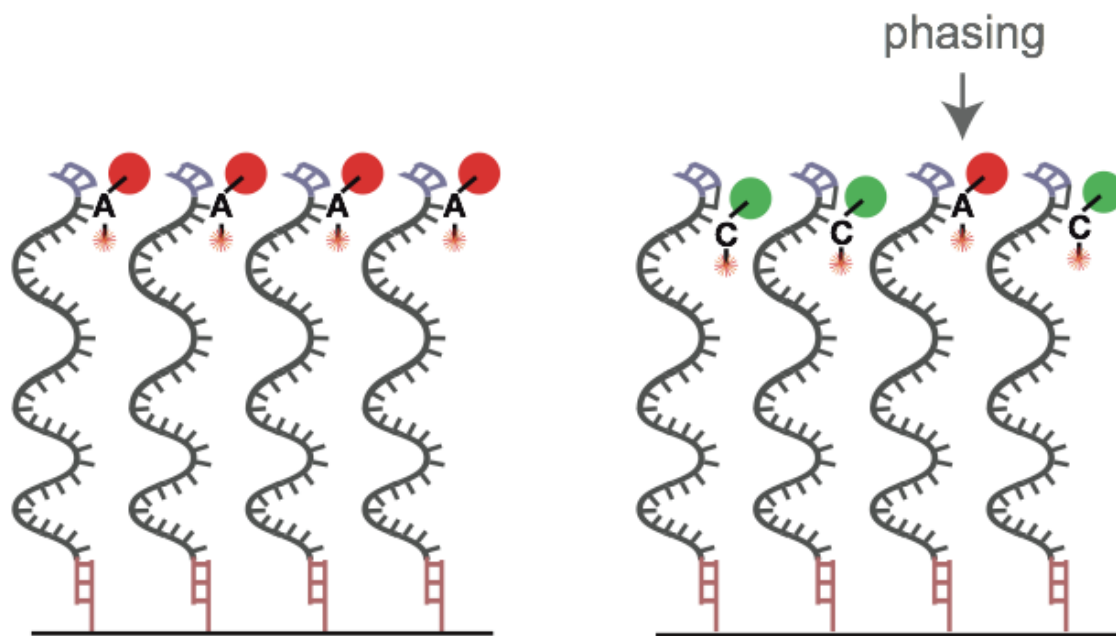
F





3. Πηγή λάθους σε Illumina

Η διαφορά φάσης στο φθορισμό λόγω ατελούς λειτουργίας των ενζύμων κοπής φθοριοφόρων/τερματιστών οδηγεί σε επιδείνωση των Phred Q κατά τη σύθεση του συμπληρωματικού κλώνου του εκμαγείου





ΤΜΗΜΑ

Βιοχημείας &
Βιοτεχνολογίας

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

3. Βασικά λάθη αλληλούχισης

- Υποκαταστάσεις για Illumina (λόγω διαφοράς φάσης) προς το τέλος του αναγνώσματος
- Ομοπολυμερή για πυροαλληλούχιση (pyrosequencing - Ion Torrent/Proton)

Ανάλυσεις προϊόντων PCR με μεθόδους αλληλούχισης νέας (NGS)

4. Ανάλυση δεδομένων



ΤΜΗΜΑ

Βιοχημείας &
Βιοτεχνολογίας
ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

4α. Προκαταρτική ανάλυση και καθαρισμός δεδομένων

- Καταμέτρηση αλληλουχιών και ανάλυση ποιοτικών χαρακτηριστικών δεδομένων
- Ποιοτικός έλεγχος αναγνωσμάτων
- Συναρμολόγηση ενισχυμάτων
- Έλεγχος για χίμαιρες και μη στόχους



ΤΜΗΜΑ

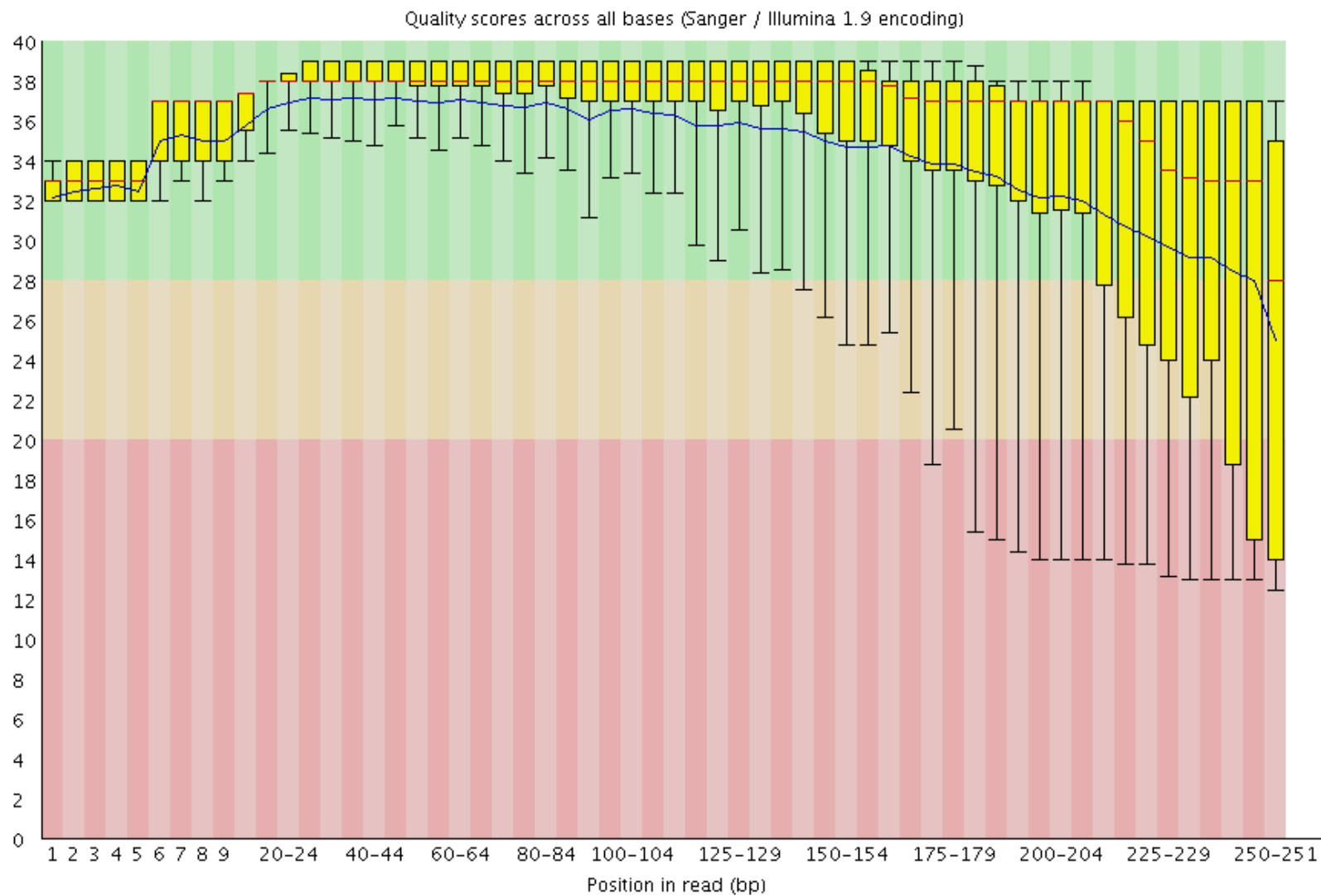
Βιοχημείας &
Βιοτεχνολογίας
ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

4α. Προκαταρτική ανάλυση και καθαρισμός δεδομένων

- Καταμέτρηση αλληλουχιών και ανάλυση ποιοτικών χαρακτηριστικών δεδομένων
- Ποιοτικός έλεγχος αναγνωσμάτων
- Συναρμολόγηση ενισχυμάτων
- Έλεγχος για χίμαιρες και μη στόχους



4α. Προκαταρτική ανάλυση και καθαρισμός δεδομένων

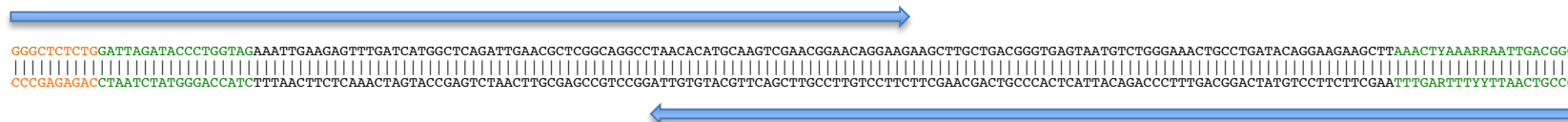




4α. Προκαταρτική ανάλυση και καθαρισμός δεδομένων

- Καταμέτρηση αλληλουχιών και ανάλυση ποιοτικών χαρακτηριστικών δεδομένων
- Ποιοτικός έλεγχος αναγνωσμάτων
- **Συναρμολόγηση ενισχυμάτων**
- Έλεγχος για χίμαιρες και μη στόχους

Read 1





4α. Προκαταρτική ανάλυση και καθαρισμός δεδομένων

- Καταμέτρηση αλληλουχιών και ανάλυση ποιοτικών χαρακτηριστικών δεδομένων
- Ποιοτικός έλεγχος αναγνωσμάτων
- Συναρμολόγηση ενισχυμάτων
- Έλεγχος για χίμαιρες και μη στόχους

sample	raw	post trimming	post assembly	Analyzed: post chimera and specificity check	Good's coverage
D1	52211	49321	48150	41947	1.000
D2	48090	45522	44260	39244	1.000
D3	30969	29485	28679	26163	0.999
M01	54529	51853	50924	45471	1.000
M02	32242	30720	30116	27542	1.000
M03	29110	27636	27096	24177	1.000
1M01	28234	26861	26331	23135	1.000
1M02	66358	62527	61334	53552	1.000
1M03	43473	41473	40668	36576	1.000
2M02	20418	19466	19063	17891	1.000
2M03	27732	26307	25777	23990	1.000
2M07	16075	15372	15065	14103	0.999
2M08	31558	29946	29423	27614	1.000
3M01	33863	32143	31605	29606	1.000
3M02	4063	3860	3789	3565	0.991
3M03	18565	17659	17320	15723	0.999
3M07	37993	36034	35466	32919	1.000
3M08	45630	43437	42745	39740	1.000
4M01	97897	93063	91498	82601	1.000
4M02	84060	80106	78854	73560	1.000
4M03	52472	49859	48920	46218	1.000
total	946358	898292	881164	803585	
remaining % of total		94.92%	93.11%	84.91%	

Read 1

GGGCTCTCTGGATTAGATACCCCTGGTAGAAATTGAAGAGTTTGTATCATGGCTCAGATTGAACGCTCGGCAGGCCTAACACATGCAAGTCGAACGGAACAGGAAGAAGCTTGCCTGACGGGTGAGTAAATGTCTGGGAACTGCCGTGATACAGGAAGAAGCTTAAACTYAAARRAATTGACGGC
CCCGAGAGACTTAATCTATGGGACCATCTTTAACTTCTCAAACCTAGTACCGAGCTTAACTTGCAGCCCTCCGGATTGTGTACGTTTCAGCTTGCCTTGTCTTCGAAAGACTGCCCACTCATTACAGACCCTTTGACGGACTATGTCTTCTTCGAATTTTGARTTTTYYTTAACTGCCC

Read 2

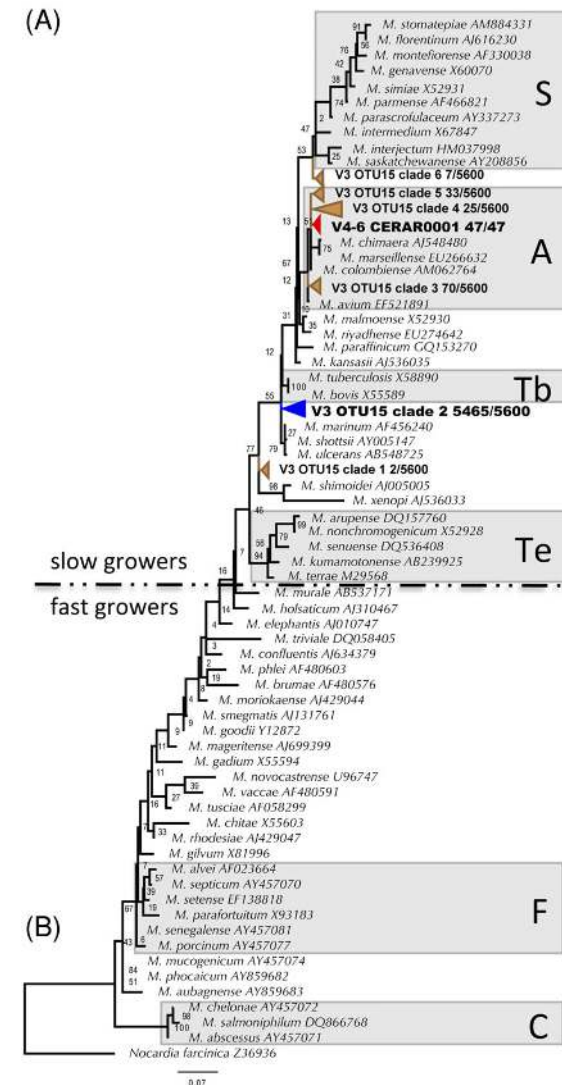


4β. Μονάδα ταξινόμησης (το δύσκολο ερώτημα)

Εξαρτάται από το ερευνητικό ερώτημα (με ποια ταξινομική βαθμίδα σχετίζεται και τι σημαίνει τη βαθμίδα αυτή για τα προς εξέταση είδη)

Συνήθεις ταξινομικές μονάδες:

- **Ταξινομικές μονάδες διαχείρισης (OTUs)** σε υπεραπλούστευση: Οι αλληλουχίες που σχηματίζουν συστάδες με το Χ% (π.χ 97% για είδος) των αλληλουχιών τους να είναι πανομοιότυπο, κατά τα πρότυπα του *Escherichia coli*.
- Ο ταξινομικός χαρακτηρισμός του κοντινότερου συγγενή σε βάση δεδομένων (**taxonomy supervised analysis**)
- Τοποθέτηση όλων των αλληλουχιών σε φυλογενετικό δένδρο και ορισμός ενός κατοφλιού (**phyloptype analysis**)





4γ. Αναλύσεις ποικιλότητας

- Ποιοτικά χαρακτηριστικά δεδομένων (πιθανότητα λάθους αναγνωσμάτων και ανασυναρμολόγησης αλληλουχίας PCR, χίμαιρες, αλληλουχίες μη στόχοι)
- **Οικολογικοί δείκτες/ερωτήματα:**
 - ❖ Δείκτες α-ποικιλότητας (σχέση ταξινομικών μονάδων εντός δείγματος)
 - ❖ Δείκτες β-ποικιλότητας (σχέση δειγμάτων εντός οικοσυστήματος)
 - ❖ Σχέση μονάδων εντός οικοσυστήματος (συσχετίσεις-συμμεταβολές / αναλύσεις δικτύου / γράφοι)
 - ❖ Σχέση ταξινομικών μονάδων με αβιοτικές παραμέτρους
- **Στατιστικά τεστ**
 - ❖ Για α-ποικιλότητα:
 - Περιγραφικές αναλύσεις όπως μέσος, διάμεσος, τυπική απόκλιση κτλ.
 - Δοκιμή υπόθεσης (ανάλυση διακύμανσης ή ANOVA, Student's t, Fisher's exact)
 - ❖ Για β-ποικιλότητα:
 - Περιγραφικές πολυμεταβλητές αναλύσεις (ανάλυση κύριων συνιστωσών ή PCA, ανάλυση αντιστοιχιών ή CA, ανάλυση κατά συστάδες, ανάλυση των κυρίων συντεταγμένων ή PCoA, μη παραμετρική πολυδιάστατη κλιμάκωση ή nMDS)
 - Δοκιμή υπόθεσης σε πολυμεταβλητές (ανάλυση πολλαπλής παλινδρόμησης ή MLRA, ανάλυση πλεονασμού ή RDA, ανάλυση κανονιστικών αντιστοιχιών ή CCA)



4γ. ποικιλότητα τύπου α (α-diversity)

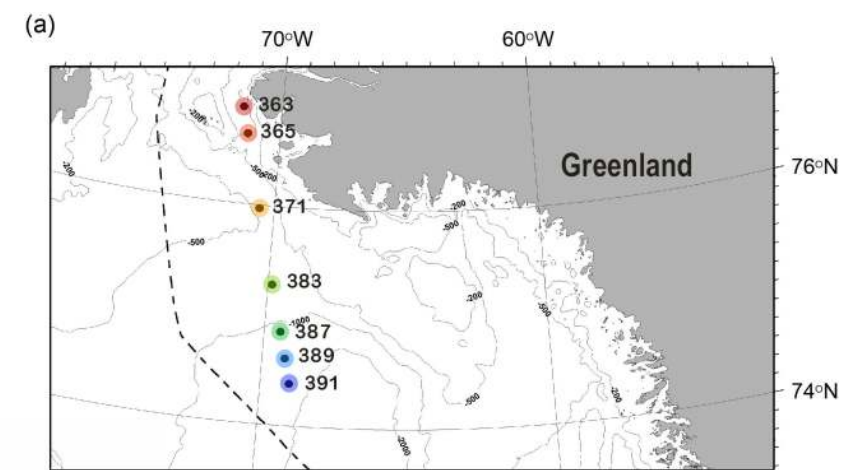
Δείκτες εντός δείγματος (παραδείγματα)

- πλούτος/πληθώρα (richness) ή εκτίμηση τους για ταξινομικές μονάδες λέγεται και ποικιλότητα μηδενικής τάξης (δε συμπεριλαμβάνει τη σχετική/απόλυτη αφθονία των μονάδων)
- δείκτης Shannon ή ποικιλότητα 1^{ης} τάξης (αντιπροσωπευτικός λιγότερο άφθονων μονάδων)
- Δείκτης Simpson ή ποικιλότητα 2^{ης} τάξης (αντιπροσωπευτικός σχετικά κυρίαρχων μονάδων)
- Δείκτης Fisher's α (ενδεικτικός της υπερκυρίαρχης ομάδας ταξινομικών μονάδων)

Ανάλυσεις προϊόντων PCR με μεθόδους αλληλούχισης νέας (NGS)

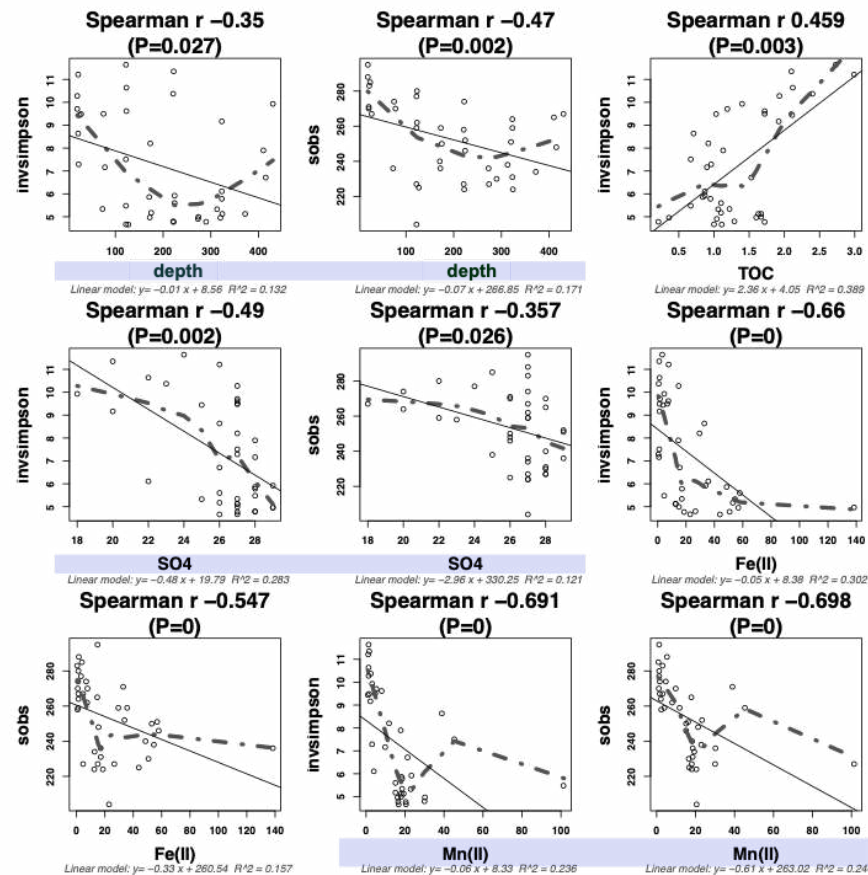
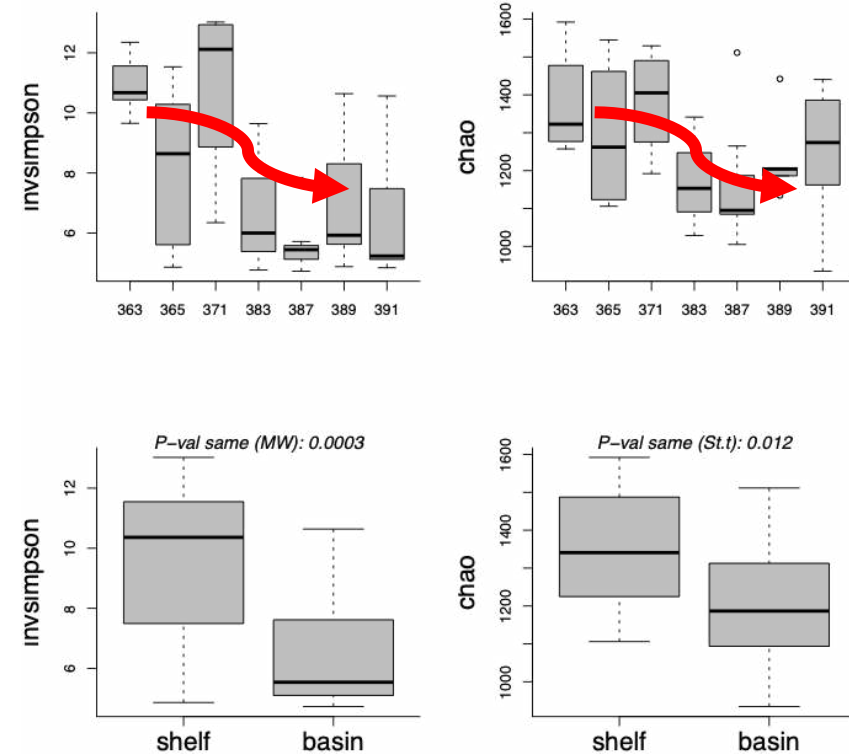
5. Ενδεικτικά αποτελέσματα

5. Ενδεικτικά αποτελέσματα (ανοιχτά της Γροιλανδίας)

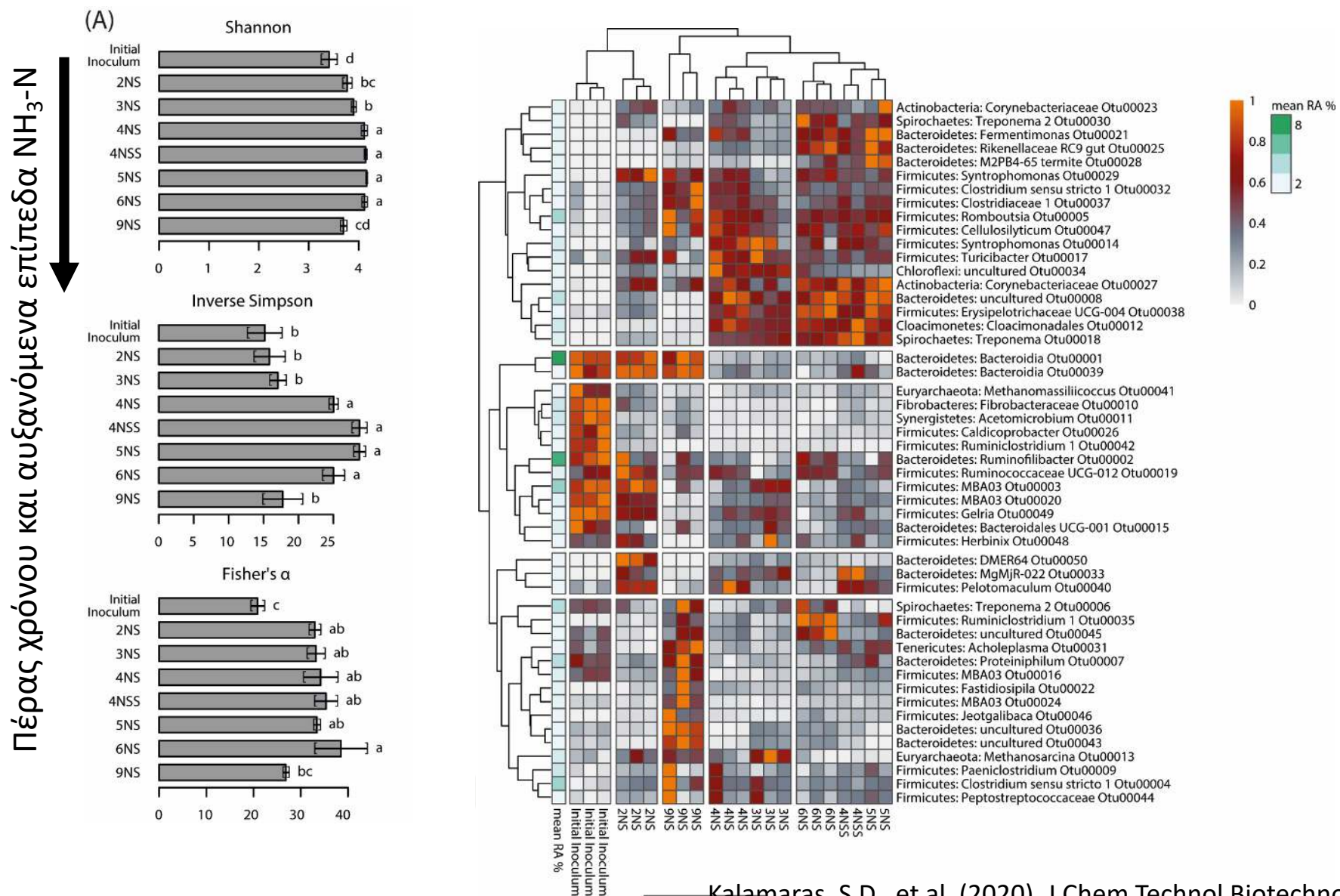


Descriptive and hypothesis testing

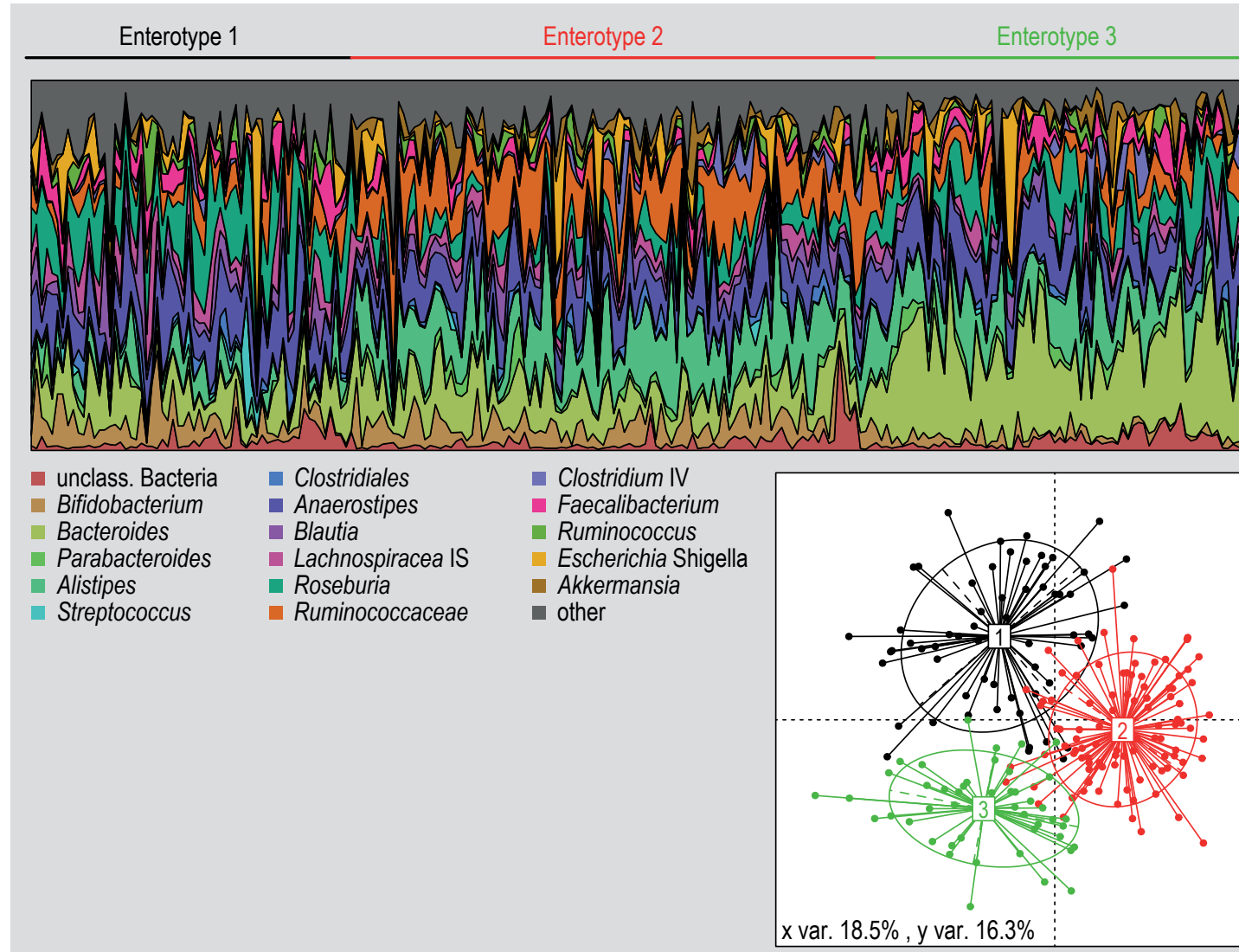
Correlation and modeling



5. Ενδεικτικά αποτελέσματα (παραγωγή βιο-CH₄ σε υψηλή συγκέντρωση NH₃)



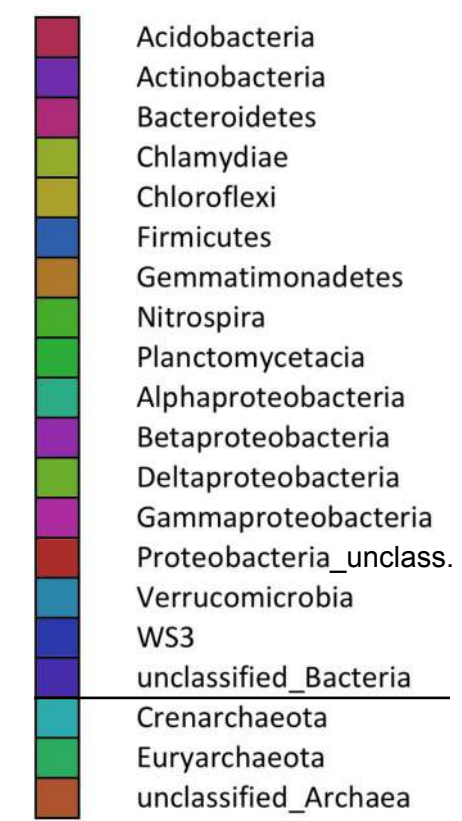
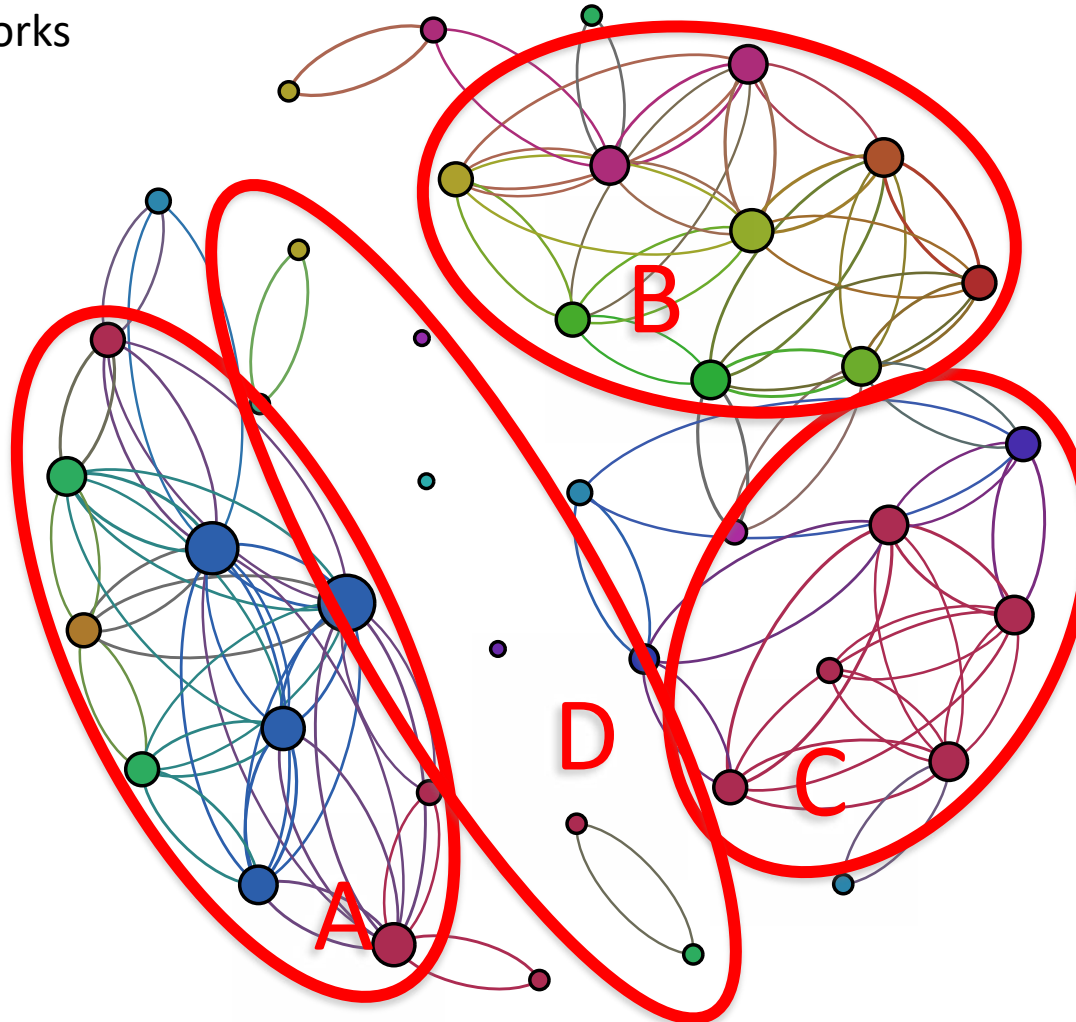
5. Ενδεικτικά αποτελέσματα (Οι «εντερότυποι» 233 παιδιών)



5. Ενδεικτικά αποτελέσματα (υποδίκτυα μικροοργανισμών σε ίζημα πηγής νερού άρδευσης στη βόρεια Ιταλία)

Microbial networks

- A**
Acidobacteria
Firmicutes
Gemmatimonadetes
Euryarchaeota
- B**
Bacteroidetes
Chlamydiae
Chloroflexi
Nitrospira
Planctomycetacia
Deltaproteobacteria
Proteobacteria unclass.
Unclassified Archaea
- C**
Acidobacteria
Gammaaproteobacteria
 WS3
Unclassified Bacteria
- D**
Acidobacteria
Actinobacteria
Chloroflexi
Alphaproteobacteria
Betaproteobacteria
Crenarchaeota
Euryarchaeota





Κάνε το και εσύ... Μπορείς!!!

- Διδακτική ενότητα του Mothur για δεδομένα Illumina (MiSeq-SOP): www.mothur.org και https://www.mothur.org/wiki/MiSeq_SOP
- Στατιστικές μέθοδοι στην οικολογία (και όχι μόνο) για ανθρώπους με το GUSTAME του Alban Ramette (<https://mb3is.megx.net/gustame>)
- Δωρεάν στατιστικά προγράμματα:
 - Χωρίς γνώσεις προγραμματισμού: PAST (<https://folk.uio.no/ohammer/past/>)
 - Γνώσεις προγραμματισμού ή καλή διάθεση: R/RStudio (<https://www.r-project.org>, <https://rstudio.com>) και προτεινόμενη μέθοδος άνευ διδασκάλου (<https://www.guru99.com/r-tutorial.html>)