



# Βιοπληροφορική

## Ενότητα 5:

Στοίχιση ακολουθιών ανά ζεύγη, 2 ΔΩ

Τμήμα: Βιοτεχνολογίας

Όνομα καθηγητή: Τ. Θηραίου



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ





# Μαθησιακοί Στόχοι

- Κατανόηση της συσχέτισης ομολογίας – ομοιότητας.
- Παρουσίαση των πληροφοριών που προκύπτουν από τη μελέτη της ομοιότητας σε επίπεδο ακολουθιών.
- Ανάλυση της στοίχισης ακολουθιών.
- Κατανόηση των διαφορετικών ειδών στοίχισης.

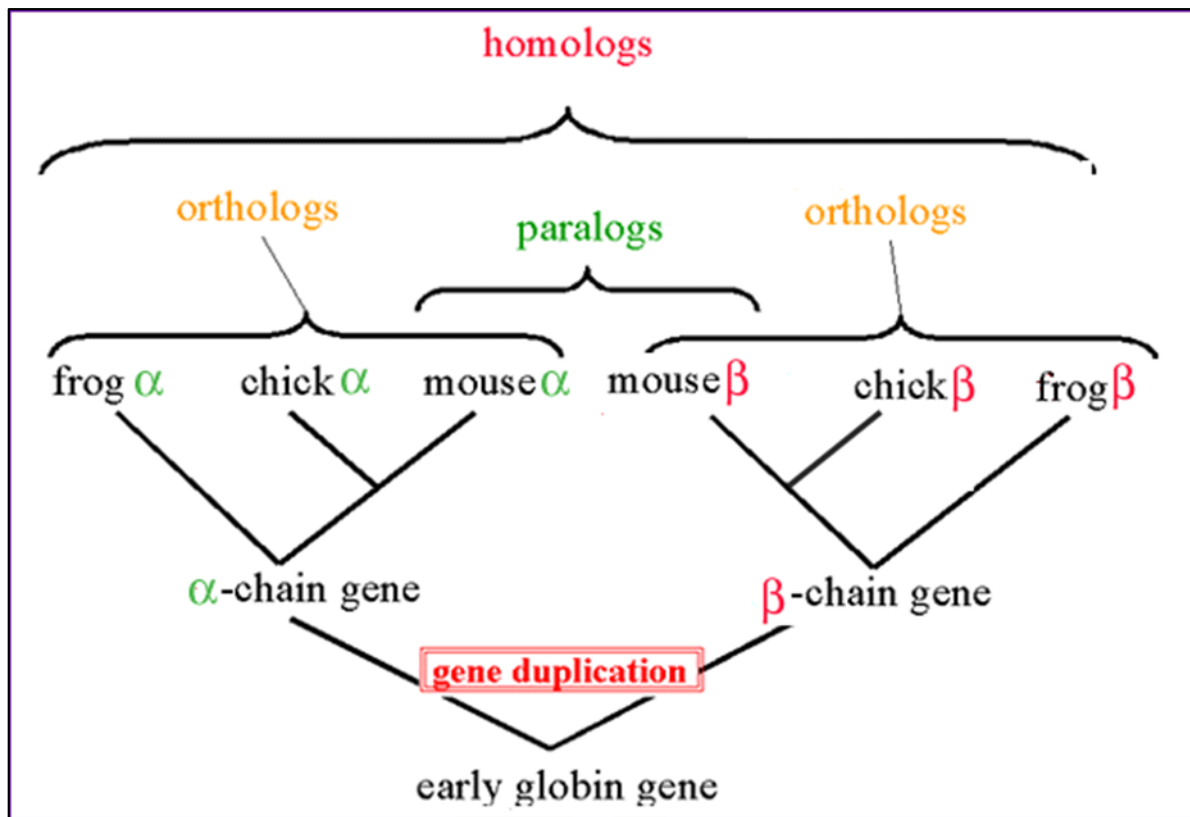


# Λέξεις Κλειδιά

- Λέξεις κλειδιά: Ομοιότητα, Ομολογία, Στοίχιση ακολουθιών ανά ζεύγη, Ολική στοίχιση ακολουθιών / Τοπική στοίχιση ακολουθιών.
- Key words: Homology, Similarity, Pairwise sequence alignment, Global sequence alignment / Local sequence alignment.



# Ομολογία 1/3





# Ομολογία 2/3

- **Ομολογία:**

- κοινή εξελικτική καταγωγή.

- **Ορθόλογα γονίδια:**

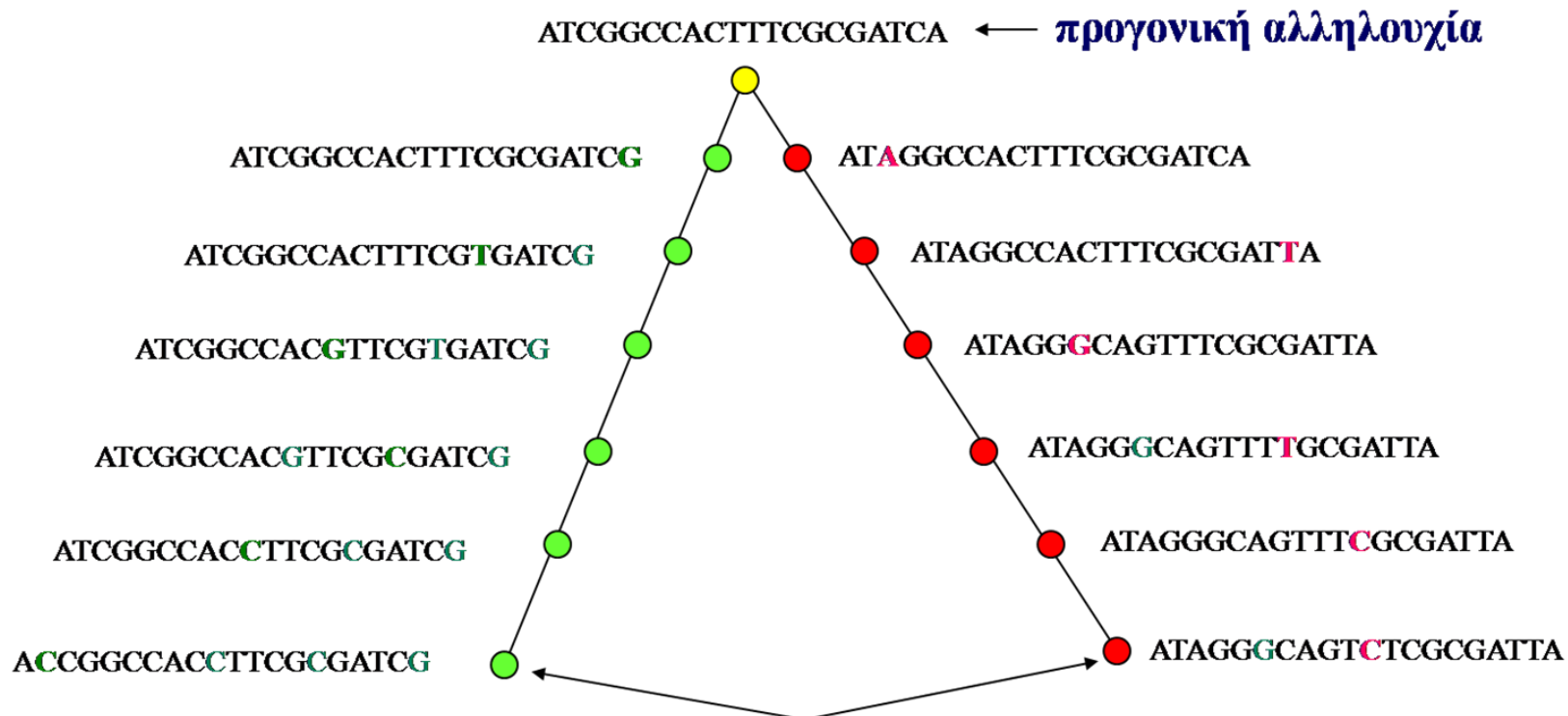
- ειδογένεση.
- συνήθως, ίδια βιολογική λειτουργία.

- **Παράλογα γονίδια:**

- γονιδιακός διπλασιασμός.
- εξελικτικό "εργαλείο" ανάπτυξης διαφοροποιημένων λειτουργιών.



# Ομολογία 3/3



Ομόλογες αλληλουχίες

ACCGGCCACCTTCGCGATCG

| | | | |

ATAGGGCAGTCTCGCGATTA



# Ομολογία / Ομοιότητα

- Ομολογία (homology)
  - κοινή εξελικτική καταγωγή.
- **Ομοιότητα (similarity)**
  - παρατηρήσιμη ποσότητα που μπορεί να εκφρασθεί ως το ποσοστό των ταυτόσημων καταλοίπων μεταξύ δύο ακολουθιών ή με τη χρήση κάποιου άλλου κατάλληλου μέτρου.
- Η **υψηλή ομοιότητα** σε επίπεδο ακολουθίας αποτελεί **ένδειξη** κοινής εξελικτικής καταγωγής - **ομολογίας**.



# Ομοιότητα

- Η μελέτη των ομοιοτήτων σε επίπεδο ακολουθίας μπορεί να δώσει πληροφορίες:
  - για την **εξελικτική σχέση** των ακολουθιών των βιομακρομορίων που παρατηρούνται στους σύγχρονους οργανισμούς.
  - για τις περιοχές των ακολουθιών οι οποίες είναι σημαντικές για τη **δομή** και τη **λειτουργία** τους.





# Μέθοδοι Σύγκρισης Ακολουθιών

- Τεχνικές που βασίζονται στη **στοίχιση** των ακολουθιών:
  - Η σύγκριση δύο ακολουθιών μπορεί να γίνει με τη στοίχισή τους και τον υπολογισμό κάποιου μέτρου ομοιότητας (ή διαφοράς) με βάση τη δεδομένη στοίχιση.

```
>P01922 | HBA_HUMAN  GSAQVKGHGKKVADALTN-----VAHVDDMPNALSALSDLHAHKL
                + +++ H  KV  +  A      V  V          L  L  +H  K
>P02240 | LGB2_LUPLU  NNPELQAHAGKVFKLVYEAAIQIQVNTGVVVTDATLKNLGSVHVS KG
```

- Τεχνικές που δεν βασίζονται στη στοίχιση των ακολουθιών.



# Στοίχιση Ακολουθιών 1/4

- απόκλιση από κοινό προγονικό μόριο μέσω:
  - **αντικαταστάσεων** (substitutions), π.χ. G στη θέση του A.
  - indels:
    - **ενθέσεων** (insertions) , π.χ. εισαγωγή του A μεταξύ των M και V.
    - **διαγραφών** (deletions) , π.χ. απαλοιφή του L μεταξύ των I και Y.

**AAVILYMV**



προγονικό μόριο

**AGVILYMV**

**AGVILYMAV**

**AGVIYMAV**

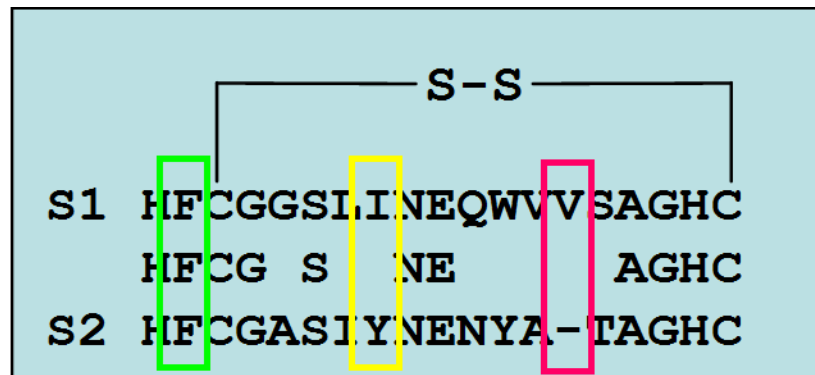


# Στοίχιση Ακολουθιών 2/4

Σε μία "ιδανική" στοίχιση:

- Η αντιστοίχιση ανόμοιων καταλοίπων (**mismatch**) αντιπροσωπεύει αντικαταστάσεις.
- Η εισαγωγή κενών (**gap**) οφείλεται σε ενθέσεις ή διαγραφές.

match



gap

mismatch



# Στοίχιση Ακολουθιών 3/4

- Δεδομένα:

- δύο ακολουθίες.
- ένα **σύστημα βαθμολόγησης** της στοίχισης:
  - όμοιων καταλοίπων (match score).
  - ανόμοιων καταλοίπων (mismatch score).
  - ενός καταλοίπου με κενό (ποινή για τα κενά, gap penalty).

- Ζητούμενο:

- εύρεση της **βέλτιστης στοίχισης** των δύο ακολουθιών που:
  - διατηρεί τη σειρά των χαρακτήρων.
  - εισάγει τα απαραίτητα κενά.
  - **μεγιστοποιεί το συνολικό score ομοιότητας.**



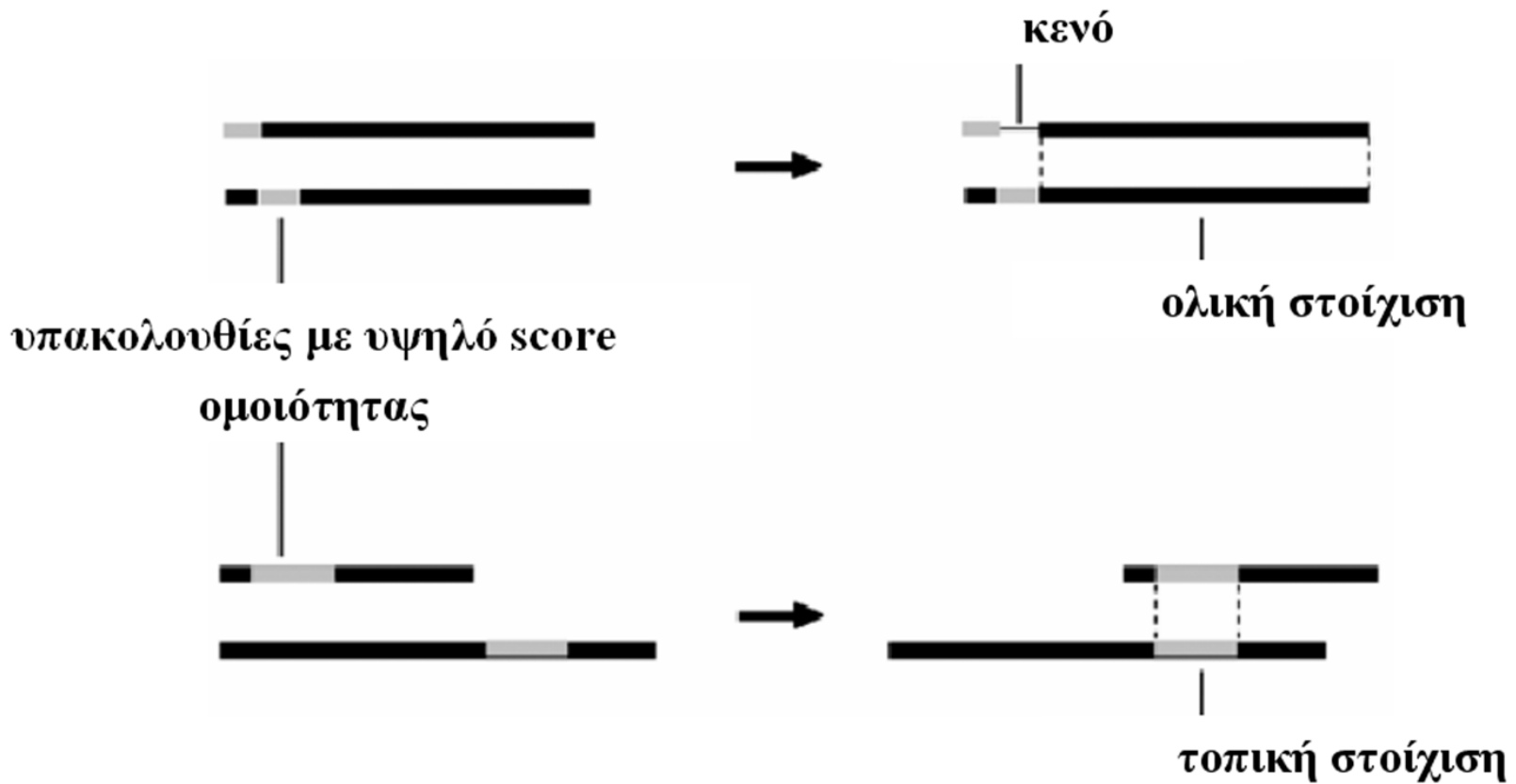
# Στοίχιση Ακολουθιών 4/4

- επιλογές

- **είδος στοίχισης.**
- **σύστημα βαθμολόγησης** (scoring system).
- **αλγόριθμος** για τον προσδιορισμό της στοίχισης.
- στατιστικές μέθοδοι για την αξιολόγηση της **σημαντικότητας της βαθμολογίας** (score) που προέκυψε κατά τη στοίχιση.
  - π.χ. είναι η ομοιότητα αρκετά υψηλή ώστε να αποτελεί ένδειξη ομολογίας???



# Είδος στοίχισης 1/3





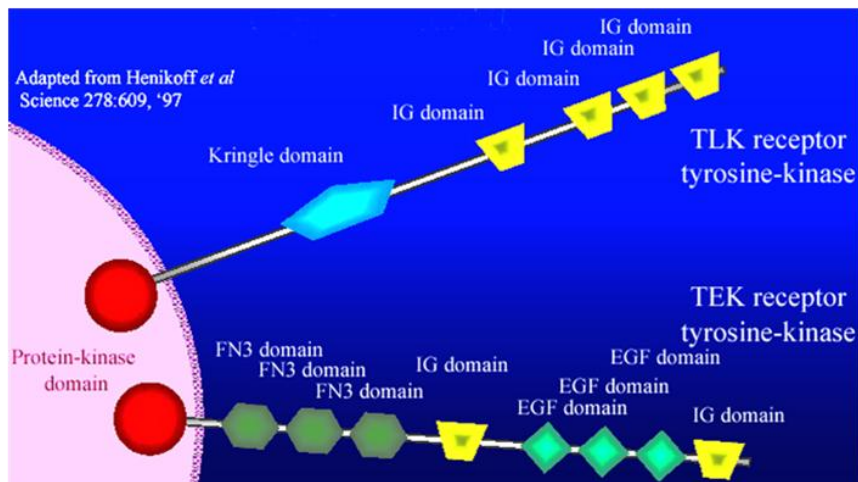
# Είδος στοίχισης 2/3

- **Ολική στοίχιση (global alignment):**
  - στοίχιση όσο το δυνατόν περισσότερων χαρακτήρων καθ' όλο το μήκος των ακολουθιών.
  - βασίζεται στο συνολικό score, ακόμα και σε βάρος τμημάτων των ακολουθιών που έχουν προφανή ομοιότητα.
  - **για κοντινές εξελικτικά ακολουθίες παρόμοιου μήκους.**



# Είδος στοίχισης 3/3

- **Τοπική στοίχιση (local alignment) :**
  - αναζήτηση όμοιων τμημάτων των ακολουθιών.
  - για ακολουθίες **διαφορετικού μήκους** με **μία ή περισσότερες συντηρημένες περιοχές**.
    - δομικά αυτοτελείς περιοχές πρωτεϊνών (domains), με χαρακτηριστική ακολουθία και λειτουργία.
    - στοίχιση mRNA με γενωμικό DNA.







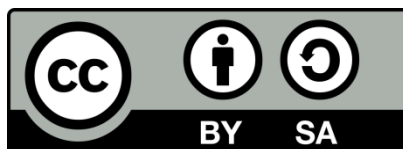
# Βιβλιογραφία

- David Mount, "Bioinformatics: Sequence and Genome Analysis", Cold Spring Harbor Laboratory Press; 2<sup>nd</sup> edition (March 12, 2013).
- Jonathan Pevsner, "Bioinformatics and Functional Genomics", Wiley-Blackwell; 2<sup>nd</sup> edition (May 4, 2009).
- Andreas D. Baxevanis, B. F. Francis Ouellette, "Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins", Wiley-Interscience; 3<sup>rd</sup> edition (October 29, 2004).



# Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδεια χρήσης, η άδεια χρήσης αναφέρεται ρητώς.





# Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα Γεωπονικού Πανεπιστημίου Αθηνών**» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ  
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ  
*επένδυση στην κοινωνία της γνώσης*

ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΣΠΑ  
2007-2013  
πρόγραμμα για την ανάπτυξη  
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ



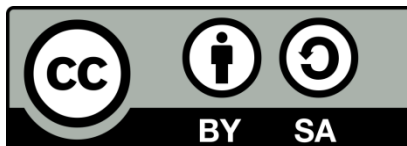
# Σημείωμα Αναφοράς

Copyright Γεωπονικό Πανεπιστήμιο Αθηνών 2015. Τμήμα Βιοτεχνολογίας, Θηραίου Τριάς. «Βιοπληροφορική». Έκδοση: 1.0. Αθήνα 2015. Διαθέσιμο από τη δικτυακή διεύθυνση:  
<https://mediasrv.aua.gr/eclass/courses/OCDB100/>



# Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Παρόμοια Διανομή 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων, π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



Η άδεια αυτή ανήκει στις άδειες που ακολουθούν τις προδιαγραφές του Ορισμού Ανοικτής Γνώσης [2], είναι ανοικτό πολιτιστικό έργο [3] και για το λόγο αυτό αποτελεί ανοικτό περιεχόμενο [4].

[1] <http://creativecommons.org/licenses/by-sa/4.0/>

[2] <http://opendefinition.org/okd/ellinika/>

[3] <http://freedomdefined.org/Definition/EI>

[4] <http://opendefinition.org/buttons/>



# Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
  - το Σημείωμα Αδειοδότησης
  - τη δήλωση Διατήρησης Σημειωμάτων
  - το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)
- μαζί με τους συνοδευόμενους υπερσυνδέσμους.