

Deep in the dark genes forest

An introduction to Metagenomics

Arthur Cousson
arthur.cousson@ird.fr



MINISTÈRE
DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE



- **Introduction**
- Metagenomics
- Applications
- Database
- Limits and Perspectives
- Conclusion

Did you say
Metagenomics ?

Introduction

- Introduction
- Metagenomics
- Applications
- Database
- Limits and Perspectives
- Conclusion

A new frontier of science is emerging that unites biology and chemistry — the exploration of natural products from previously uncultured soil microorganisms. The approach involves directly accessing the genomes of soil organisms that cannot be, or have not been, cultured by isolating their DNA, cloning it into culturable organisms and screening the resultant clones for the production of new chemicals. The excitement surrounding this new field lies in the vast diversity of unknown soil microflora and the chemical richness that they are thought to contain. The methodology has been made possible by advances in molecular biology and eukaryotic genomics, which have laid the groundwork for cloning and functional analysis of the collective genomes of soil microflora, which we term the metagenome of the soil.

Metagenomics

- Introduction
- Metagenomics
- Applications
- Database
- Limits and Perspectives
- Conclusion

Tree of Life Problem



Scala naturae
Didacus Valades
Rhetorica Christiana
1579

Metagenomics

- Introduction
- Metagenomics
- Applications
- Database
- Limits and Perspectives
- Conclusion

Tree of Life Problem

IDEE D'UNE ECHELLE

DES ETRES NATURELS.

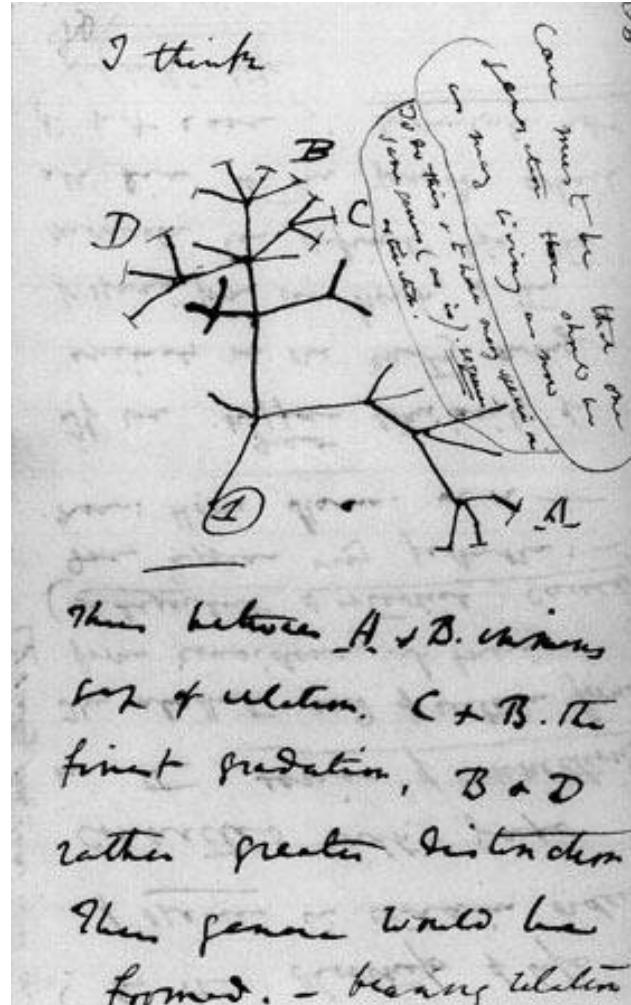
L'HOMME.	Limaçons.	Lithophytes.
Orang-Outang.	COQUILLAGES.	Amianthe.
Singe.	Vers à tuyau.	Talcs, Gyps, Sélénites.
QUADRUPEDES.	Teignes.	Ardoises.
Ecureuil volant.	INSECTES.	PIERRES.
Chauveouris.	Gallinectes.	Pierres figurées.
Autruche.	Tent., ou Solitaire.	Crystallisations.
OISEAUX.	Polypes.	SEL.S.
Oiseaux aquatiques.	Orties de Mcr.	Vitriols.
Oiseaux amphibiens.	Sensitive.	METAUX.
Poissons volans.	PLANTES.	DEMI-METAUX.
POISSONS.	Lychens.	SOUFRES.
Poissons rampans.	Moisières.	Bitumes.
Anguilles.	Champignons, Agarics.	TERRES.
Serpents d'eau.	Truffes.	Terre pure.
SERPENS.	Coraux & Coralloïdes.	EAU.
Limaces.	Lithophytes.	AIR.
Limaçons.	Amianthe.	FEU.
COQUILLAGES.	Talcs, Gyps, Sélénites.	Matieres plus subtiles.
Vers à tuyau.	Ardoises.	
	PIERRES.	

Charles Bonnet
Traité
d'insectologie,
1745

Metagenomics

- Introduction
- Metagenomics
- Applications
- Database
- Limits and Perspectives
- Conclusion

Tree of Life Problem

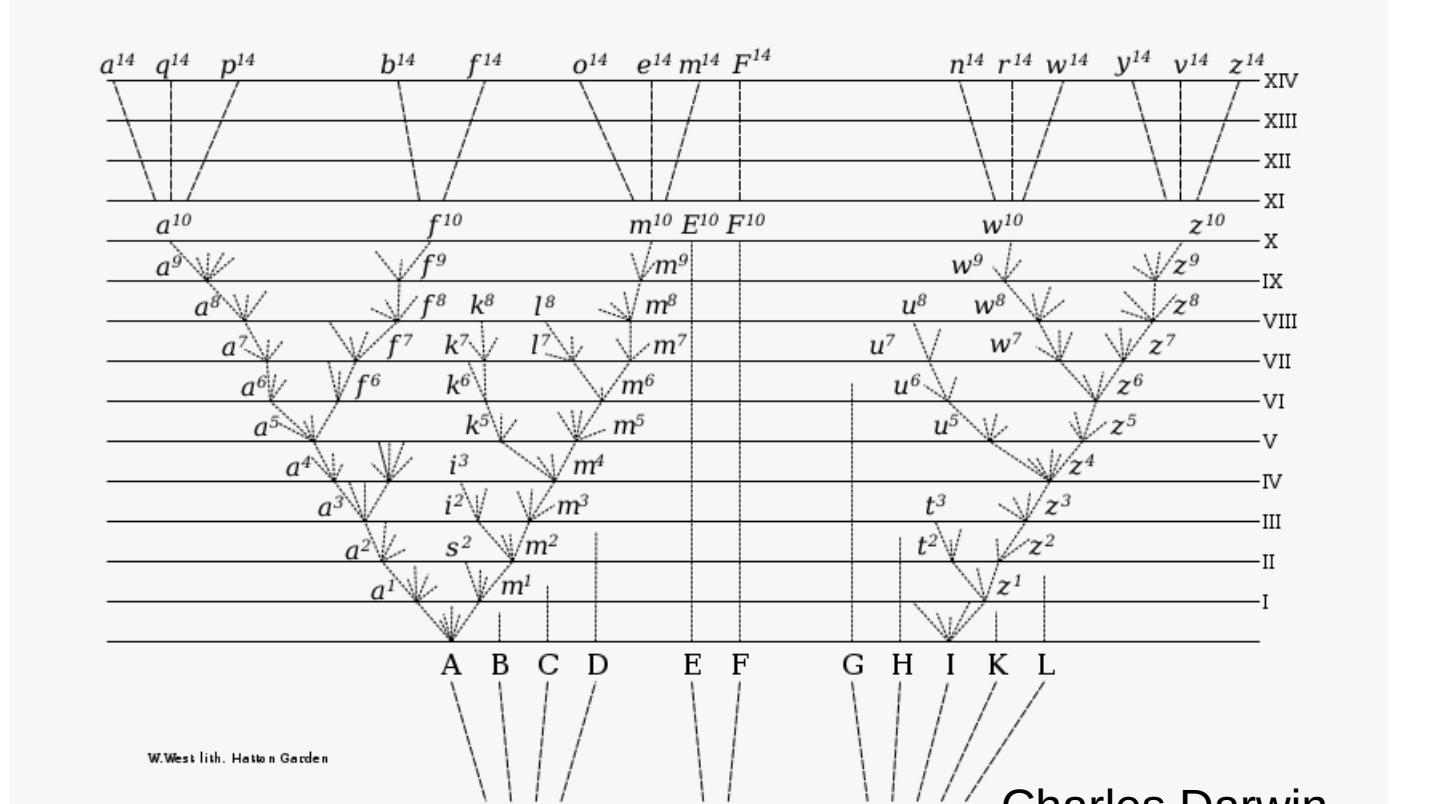


Charles Darwin
Personal notes
1837

Metagenomics

- Introduction
- Metagenomics
- Applications
- Database
- Limits and Perspectives
- Conclusion

Tree of Life Problem

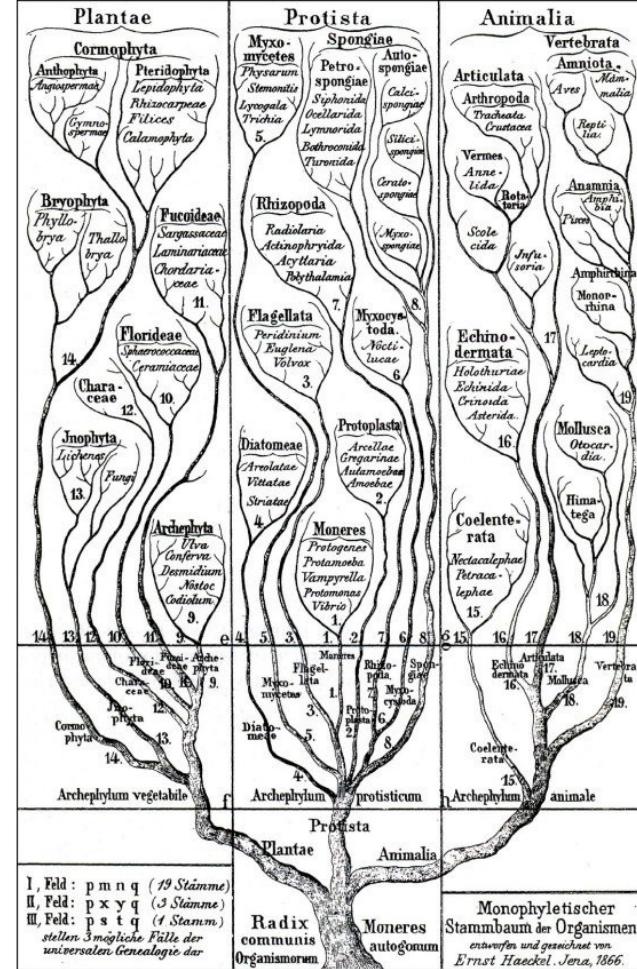


Charles Darwin
On the Origin of Species
1859

Metagenomics

- Introduction
- Metagenomics
- Applications
- Database
- Limits and Perspectives
- Conclusion

Tree of Life Problem

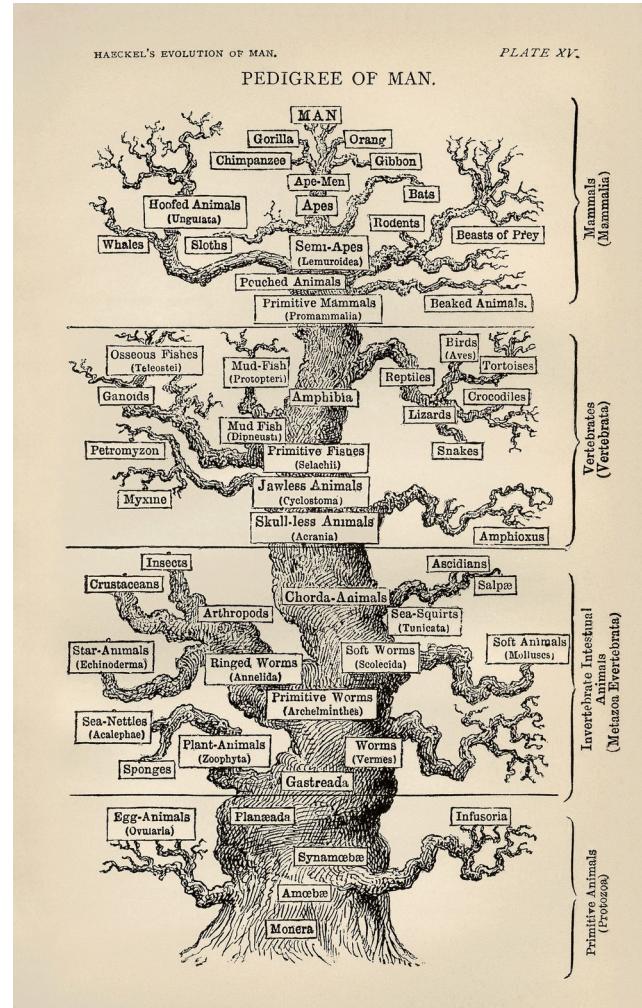


Ernst Haeckel
1866

Metagenomics

- Introduction
- Metagenomics
- Applications
- Database
- Limits and Perspectives
- Conclusion

Tree of Life Problem



Ernst Haeckel
1886

Metagenomics

- Introduction
- Metagenomics
- Applications
- Database
- Limits and Perspectives
- Conclusion

Tree of Life Problem

4578 Evolution: Woese *et al.*

Proc. Natl. Acad. Sci. USA 87 (1990)

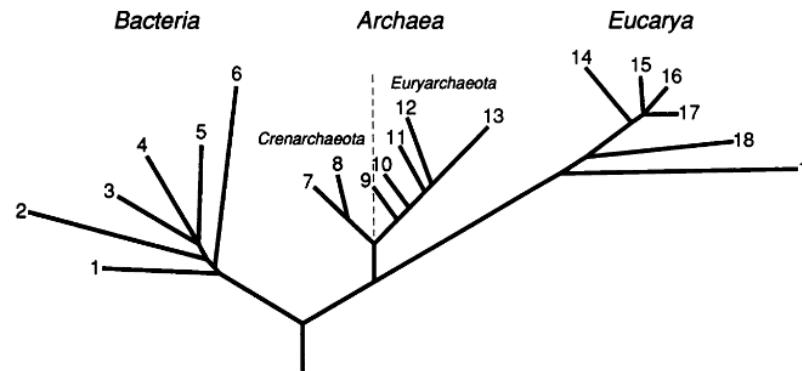


FIG. 1. Universal phylogenetic tree in rooted form, showing the three domains. Branching order and branch lengths are based upon rRNA sequence comparisons (and have been taken from figure 4 of ref. 2). The position of the root was determined by comparing (the few known) sequences of pairs of paralogous genes that diverged from each other before the three primary lineages emerged from their common ancestral condition (27). [This rooting strategy (28) in effect uses the one set of (aboriginally duplicated) genes as an outgroup for the other.] The numbers on the branch tips correspond to the following groups of organisms (2). Bacteria: 1, the Thermotogales; 2, the flavobacteria and relatives; 3, the cyanobacteria; 4, the purple bacteria; 5, the Gram-positive bacteria; and 6, the green nonsulfur bacteria. Archae: the kingdom Crenarchaeota: 7, the genus *Pyrodictium*; and 8, the genus *Thermoproteus*; and the kingdom Euryarchaeota: 9, the Thermococcales; 10, the Methanococcales; 11, the Methanobacteriales; 12, the Methanomicrobiales; and 13, the extreme halophiles. Eucarya: 14, the animals; 15, the ciliates; 16, the green plants; 17, the fungi; 18, the flagellates; and 19, the microsporidia.

Metagenomics

- Introduction
- Metagenomics
- Applications
- Database
- Limits and Perspectives
- Conclusion

Tree of Life Problem

4578 Evolution: Woese *et al.*

Proc. Natl. Acad. Sci. USA 87 (1990)

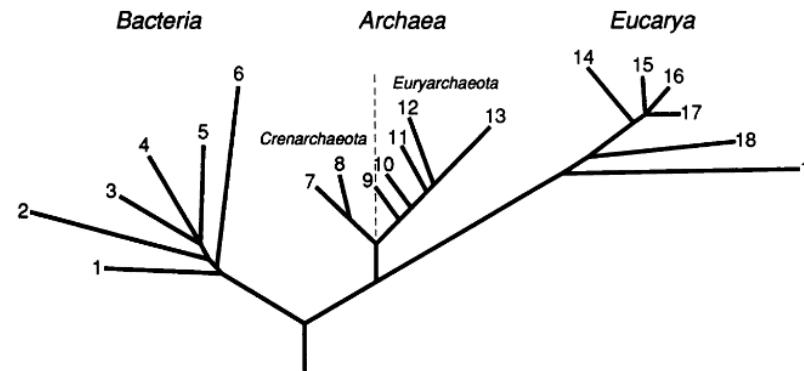


FIG. 1. Universal phylogenetic tree in rooted form, showing the three domains. Branching order and branch lengths are based upon rRNA sequence comparisons (and have been taken from figure 4 of ref. 2). The position of the root was determined by comparing (the few known) sequences of pairs of paralogous genes that diverged from each other before the three primary lineages emerged from their common ancestral condition (27). [This rooting strategy (28) in effect uses the one set of (aboriginally duplicated) genes as an outgroup for the other.] The numbers on the branch tips correspond to the following groups of organisms (2). Bacteria: 1, the Thermotogales; 2, the flavobacteria and relatives; 3, the cyanobacteria; 4, the purple bacteria; 5, the Gram-positive bacteria; and 6, the green nonsulfur bacteria. Archae: the kingdom Crenarchaeota: 7, the genus *Pyrodictium*; and 8, the genus *Thermoproteus*; and the kingdom Euryarchaeota: 9, the Thermococcales; 10, the Methanococcales; 11, the Methanobacteriales; 12, the Methanomicrobiales; and 13, the extreme halophiles. Eucarya: 14, the animals; 15, the ciliates; 16, the green plants; 17, the fungi; 18, the flagellates; and 19, the microsporidia.

Woese basically invented the Archaea through barcoding

Tree of Life Problem

- Introduction
- Metagenomics
- Applications
- Database
- Limits and Perspectives
- Conclusion

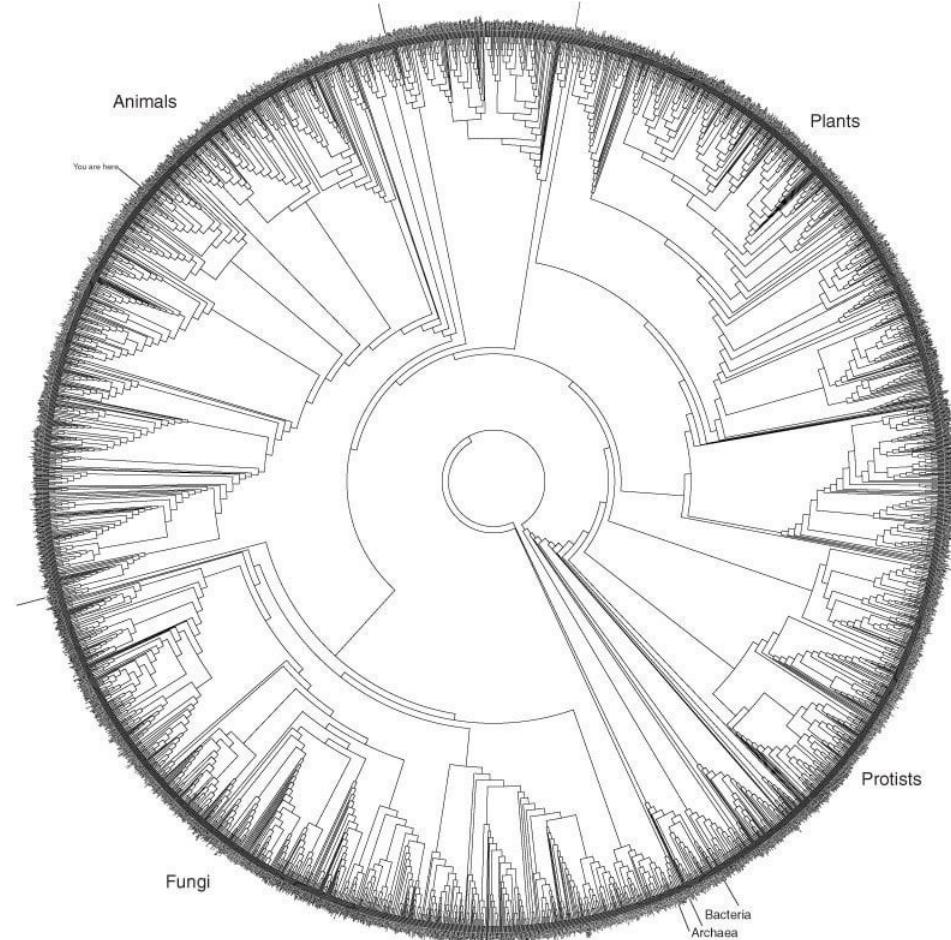
“The cell is basically a historical document, and gaining the capacity to read it (by sequencing of genes) cannot but drastically alter the way we look at all biology.”

Woese, 1987

Metagenomics

- **Introduction**
- Metagenomics
- Applications
- Database
- Limits and Perspectives
- Conclusion

Tree of Life Problem

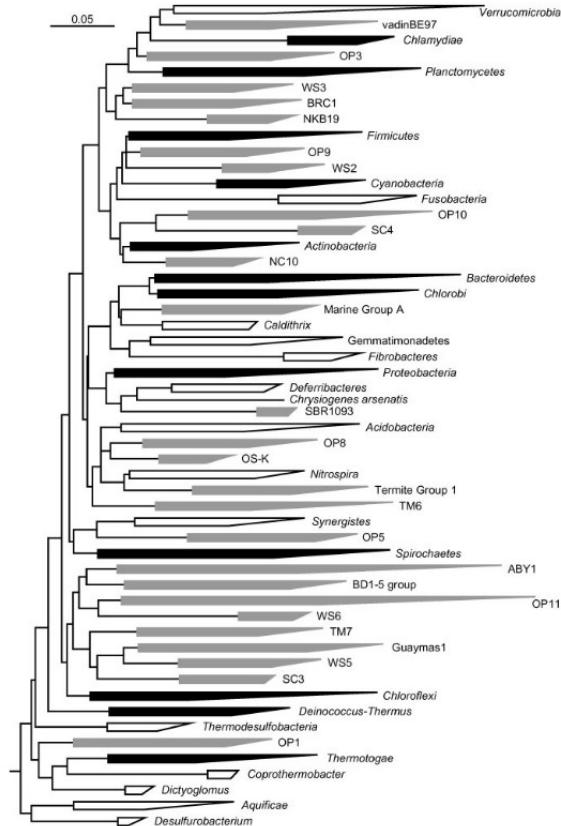


Metagenomics

- Introduction
- Metagenomics
- Applications
- Database
- Limits and Perspectives
- Conclusion

Tree of Life Problem

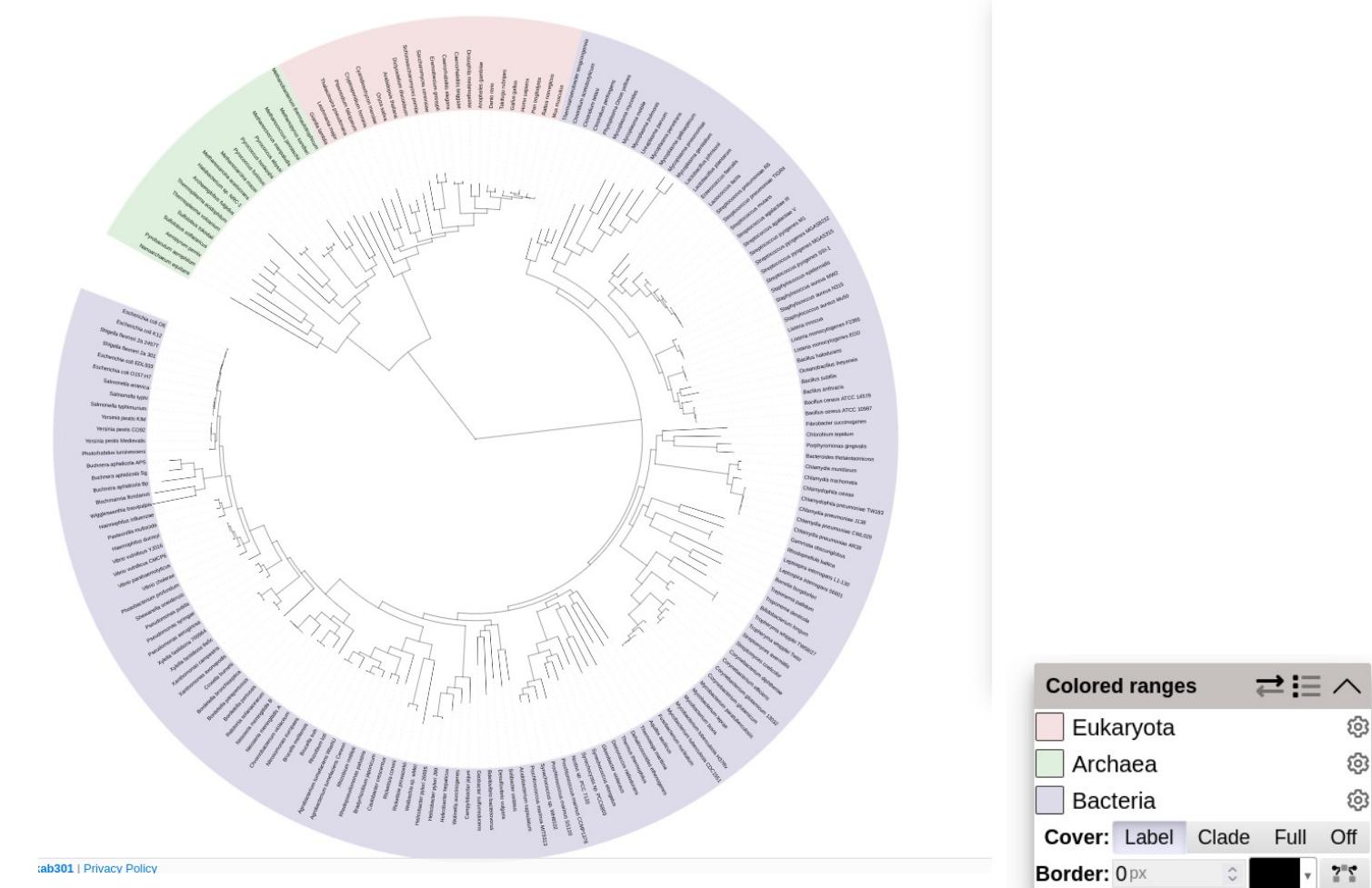
374 RAPPÉ ■ GIOVANNONI

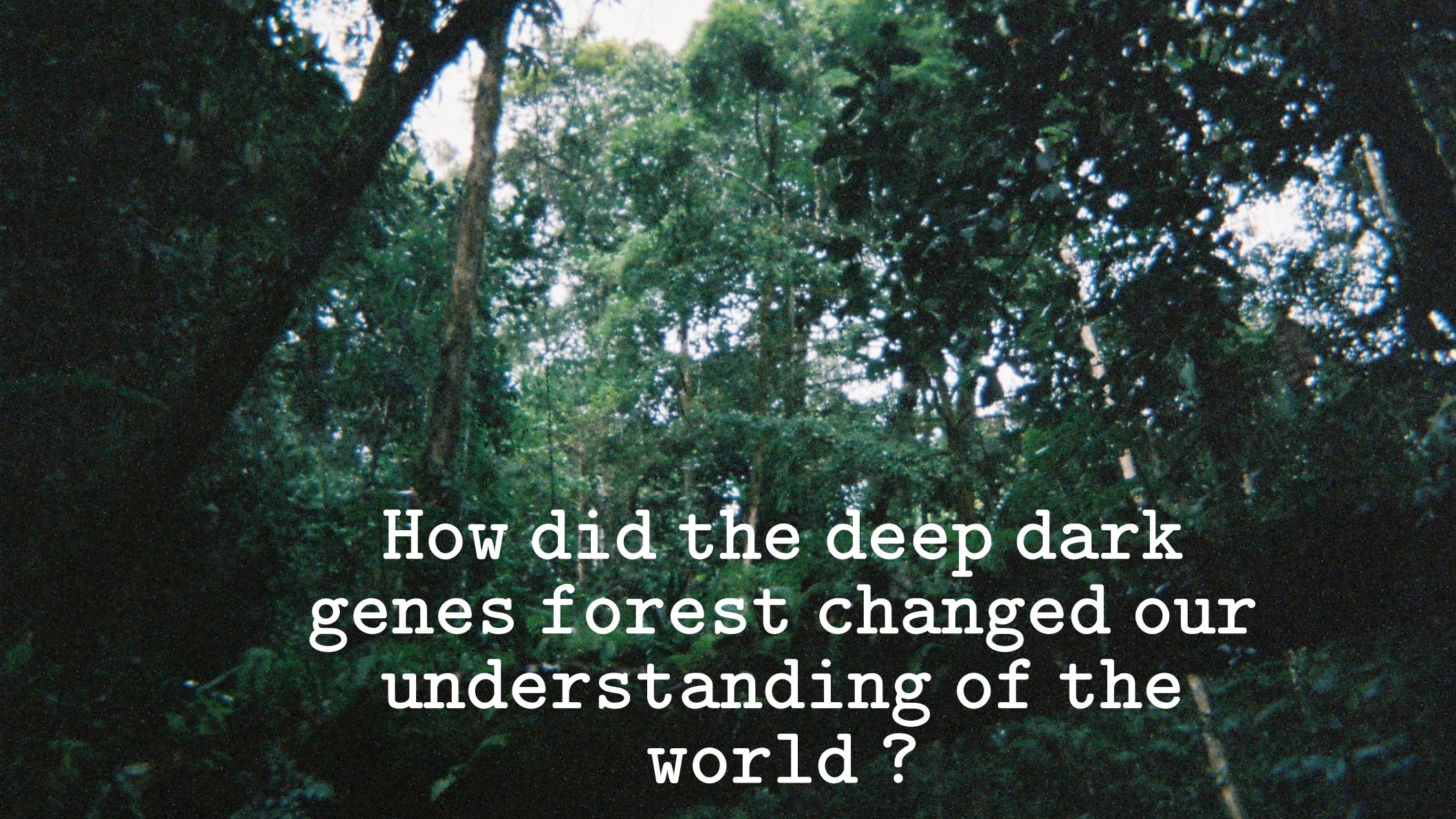


Metagenomics

- Introduction
- Metagenomics
- Applications
- Database
- Limits and Perspectives
- Conclusion

Tree of Life Problem



A photograph of a dense forest. The foreground is dominated by dark green foliage and branches. In the background, tall trees rise against a bright sky, with sunlight filtering through the leaves in several bright, starburst-like patterns.

How did the deep dark
genes forest changed our
understanding of the
world ?

Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Origins



Bacteria, again and again

Metagenomics

I – Metagenomics : Origins

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

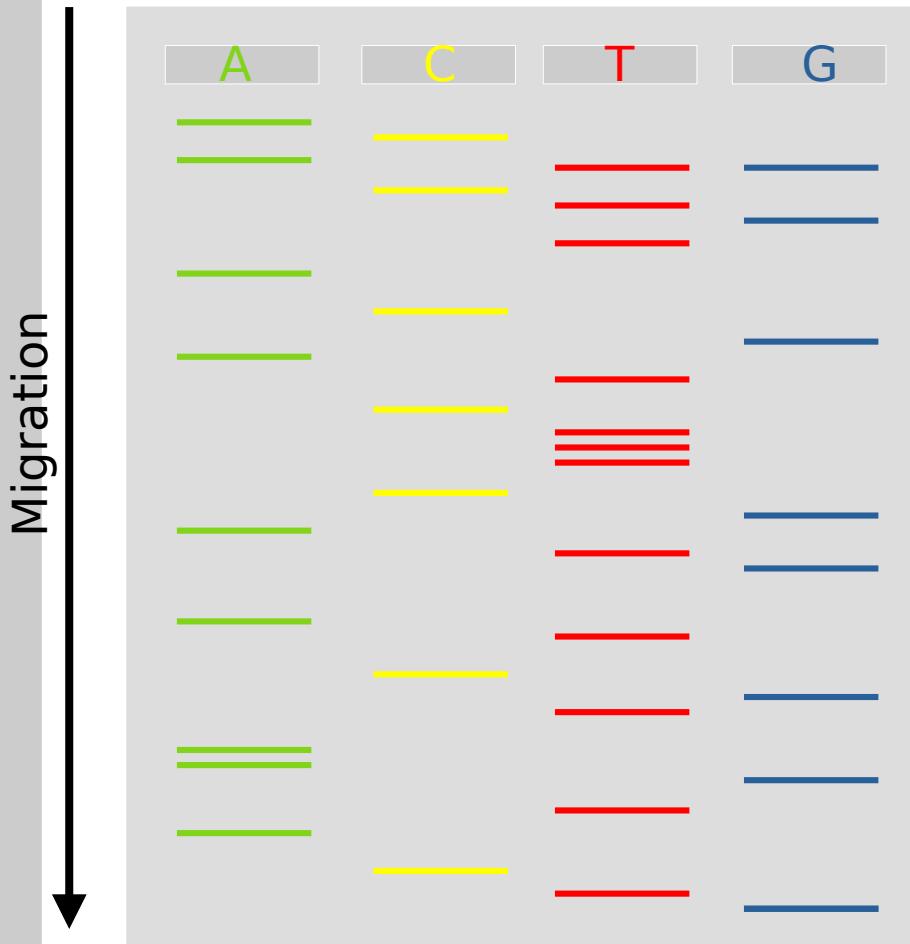


Around 30 bacterial phyla described and cultivated

Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Origins



First Generation
DNA Sequencing
Sanger
Sequencing
1970s

Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Origins

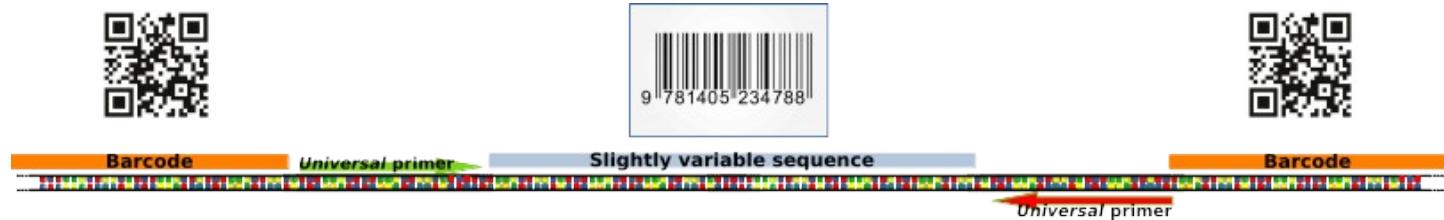
Barcode



- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Origins

Metabarcoding



- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

Construction of Phylogenetic Trees

A method based on mutation distances as estimated from cytochrome *c* sequences is of general applicability.

Walter M. Fitch and Emanuel Margoliash

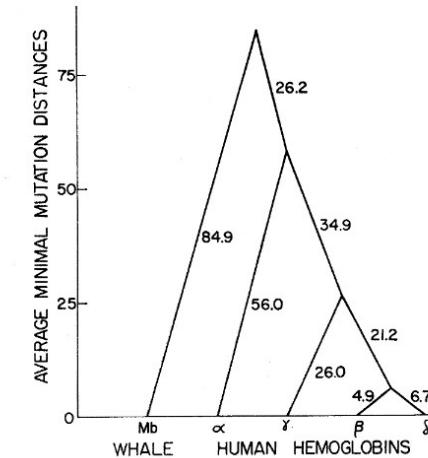
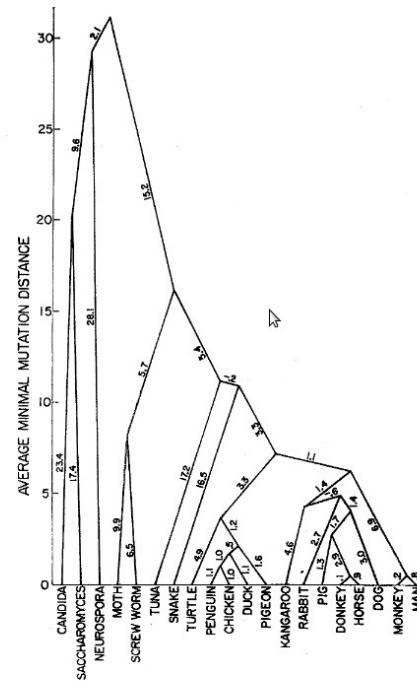


Fig. 2 (left). Phylogeny as reconstructed from observable mutations in the cytochrome *c* gene. Each number on the figure is the corrected mutation distance (see text) along the line of descent as determined from the best computer fit so far found. Each apex is placed at an ordinate value representing the average of the sums of all mutations in the lines of descent from that apex.

Fig. 3 (right above). A gene phylogeny as reconstructed from observable mutations in several heme-containing globins. See Fig. 2 for details. The percent "standard deviation" (7) for this tree is 1.33.

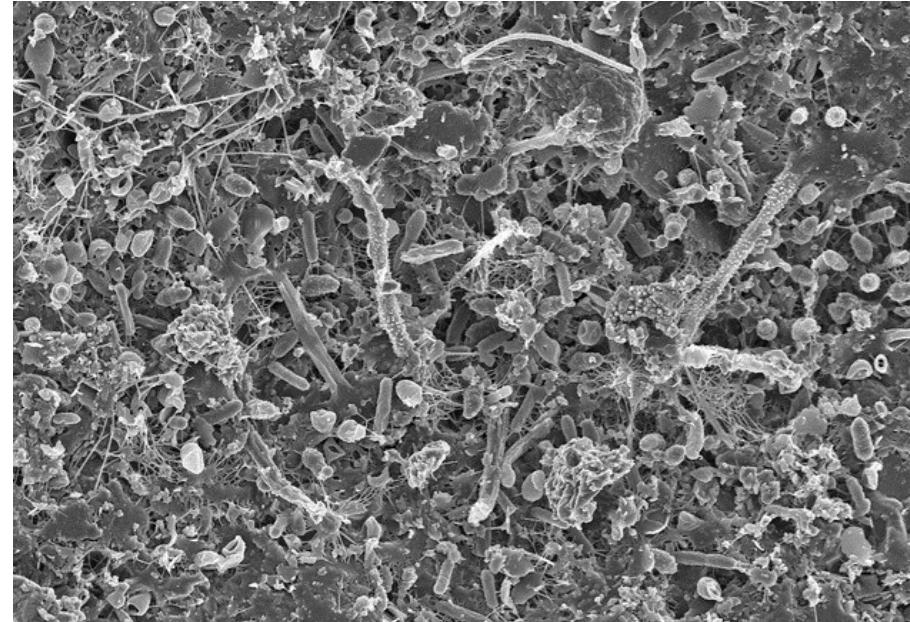
The background of the image is a dark, atmospheric landscape. It features a dense, dark green foreground that gradually transitions into a lighter, misty area. In the distance, there are faint outlines of mountain peaks under a sky filled with heavy, greyish-white clouds.

How would you enter in the
forest ?

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Shotgun Metagenomics

The first modern metagenomics discoveries



Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Shotgun Metagenomics

The first modern metagenomics discoveries



Rondon et al 2000

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Shotgun Metagenomics

The first modern metagenomics discoveries

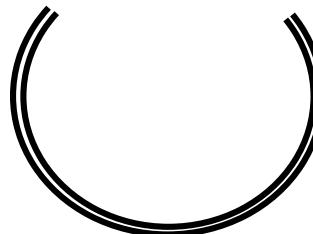


DNA Fragment

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Shotgun Metagenomics

The first modern metagenomics discoveries

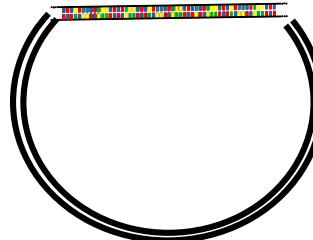


Opened plasmid

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Shotgun Metagenomics

The first modern metagenomics discoveries



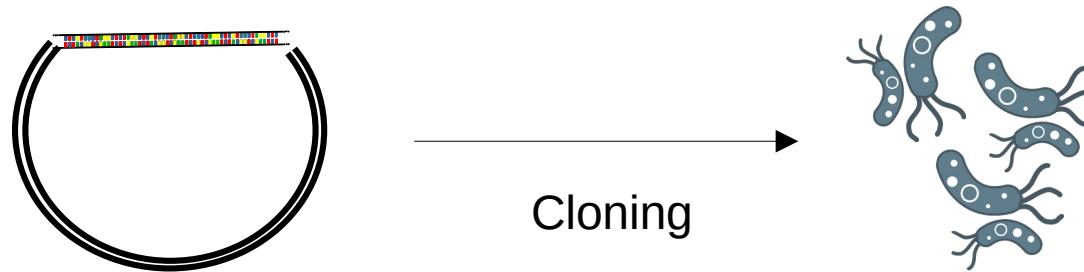
Bacterial Artificial Chromosome
BAC

Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Shotgun Metagenomics

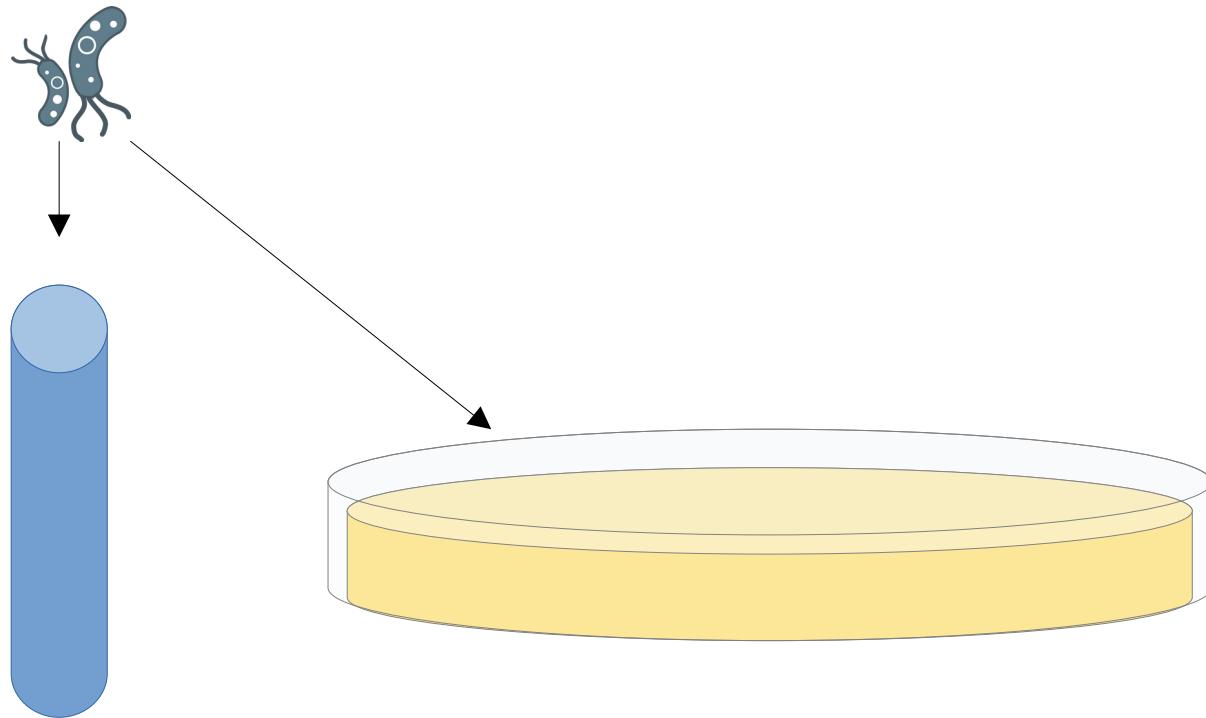
The first modern metagenomics discoveries



- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Shotgun Metagenomics

The first modern metagenomics discoveries



- Introduction
- Metagenomics
 - Origins
 - Methodologies
 - Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

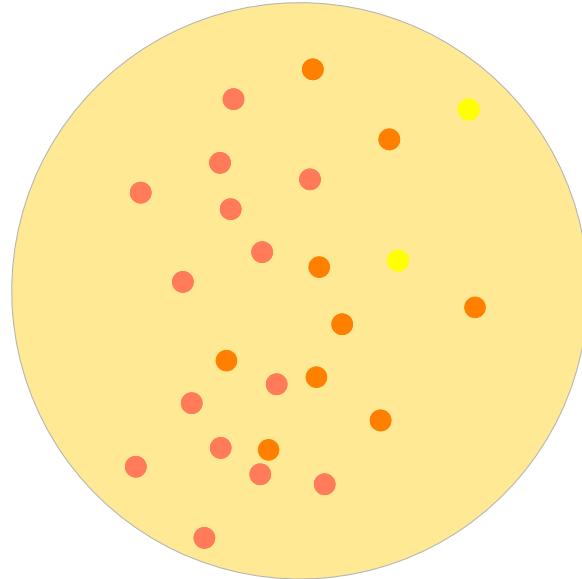
I – Metagenomics : Shotgun Metagenomics

- Testing enzymatic capacities
 - Antibiotic resistance
 - Color
 - Smell
 - Taste
 - Toxicity
 - Allelopathy
 - Biofilm formation
 - Growth rate

Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

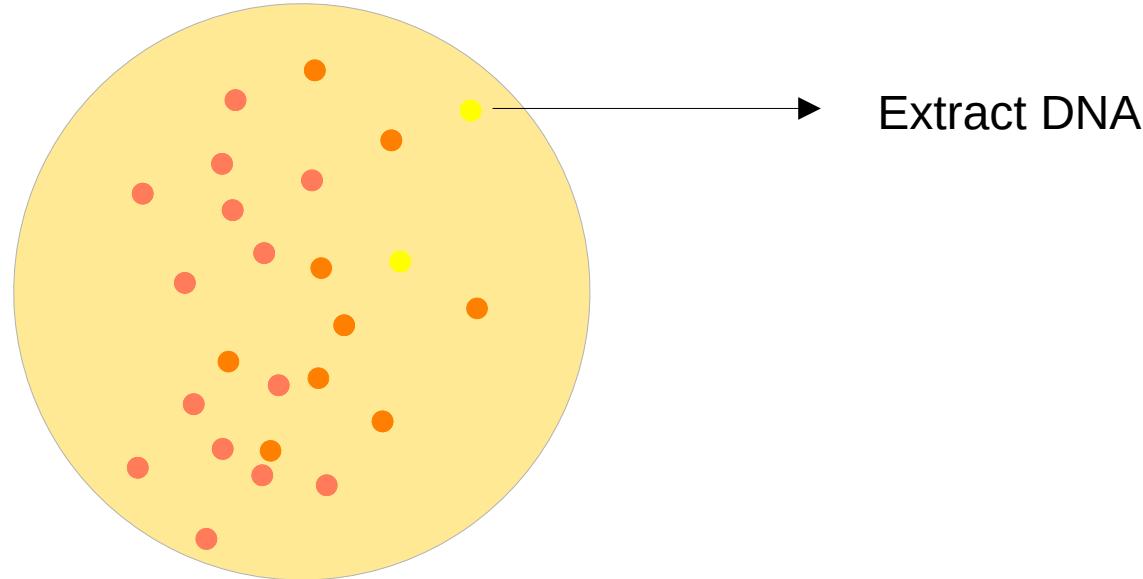
I – Metagenomics : Shotgun Metagenomics



Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

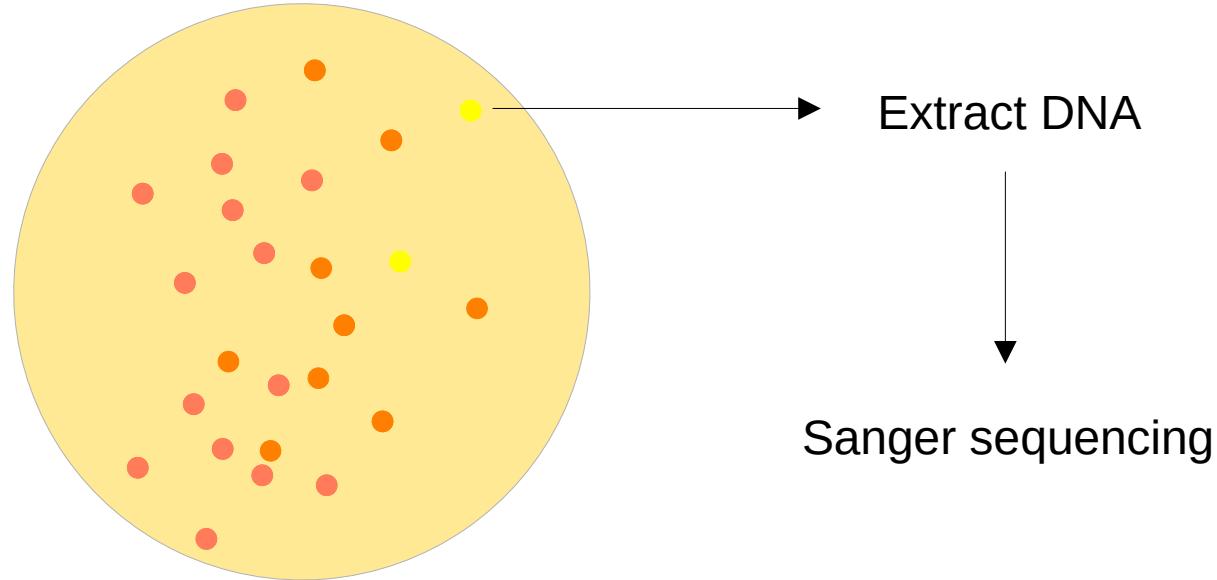
I – Metagenomics : Shotgun Metagenomics



Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

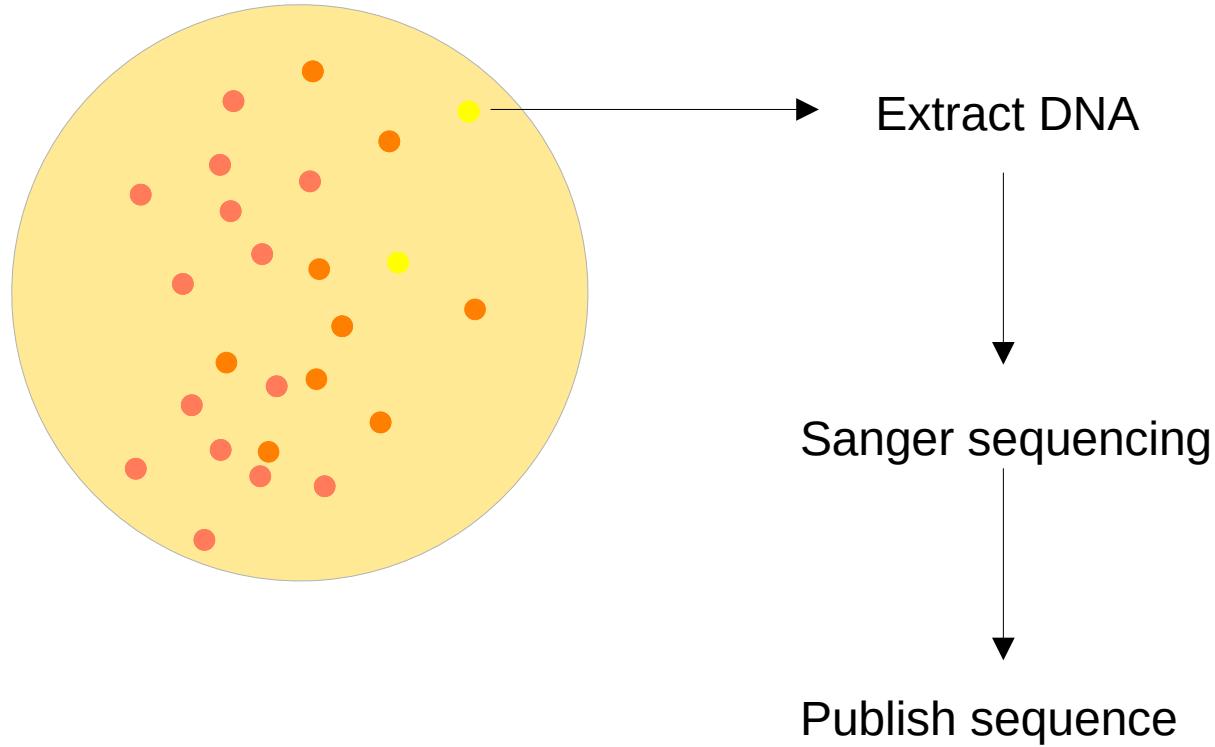
I – Metagenomics : Shotgun Metagenomics



Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

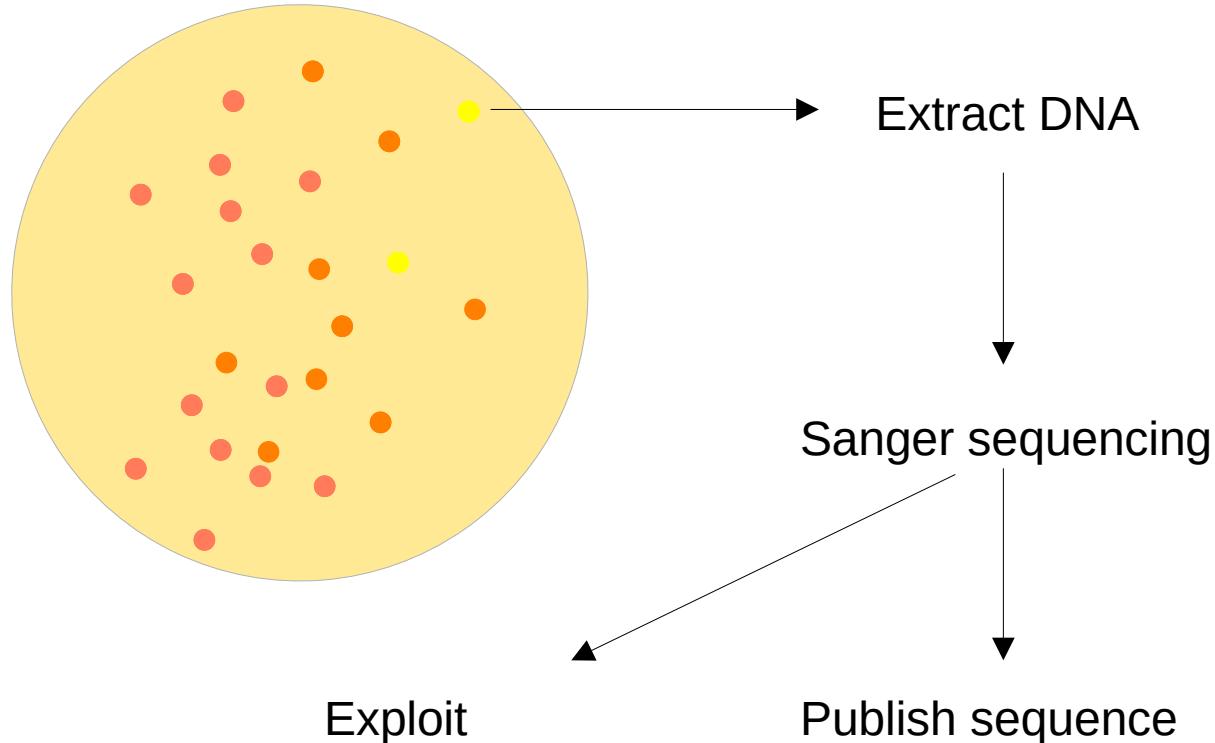
I – Metagenomics : Shotgun Metagenomics

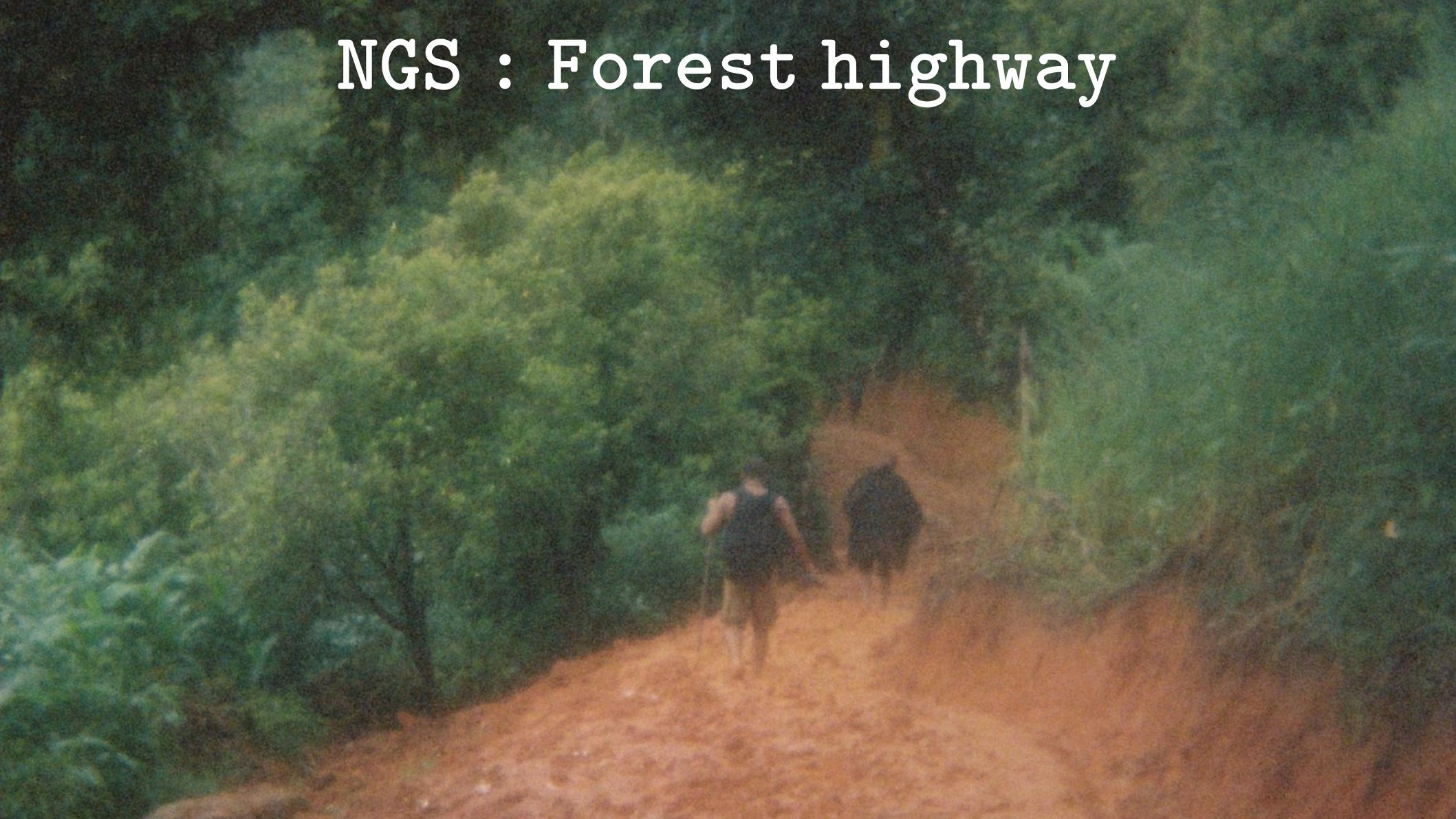


Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Shotgun Metagenomics



A photograph showing a person walking away from the camera on a dirt path through a dense forest. The path is reddish-brown and curves to the right. The surrounding trees are tall and green, with sunlight filtering through the canopy. The person is wearing dark clothing and a backpack.

NGS : Forest highway

- Introduction
- Metagenomics
 - Origins
 - Methodologies
 - Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

1) Select an ecosystem

- Introduction
- Metagenomics
 - Origins
 - Methodologies
 - Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

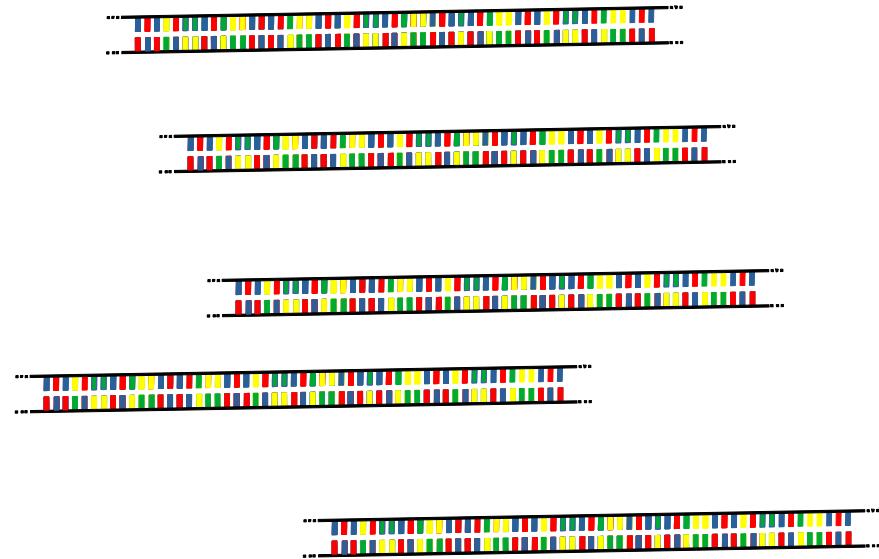
1) Select an ecosystem



- Introduction
- Metagenomics
 - Origins
 - Methodologies
 - Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Modern Methodologies

- 1) Select an ecosystem
- 2) Extract DNA

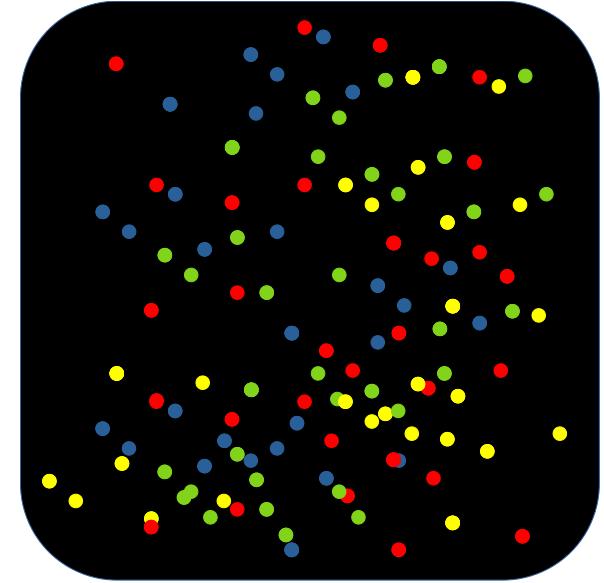


Metagenomics

- Introduction
- Metagenomics
 - Origins
 - Methodologies
 - Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Modern Methodologies

- 1) Select an ecosystem
- 2) Extract DNA
- 3) Sequence

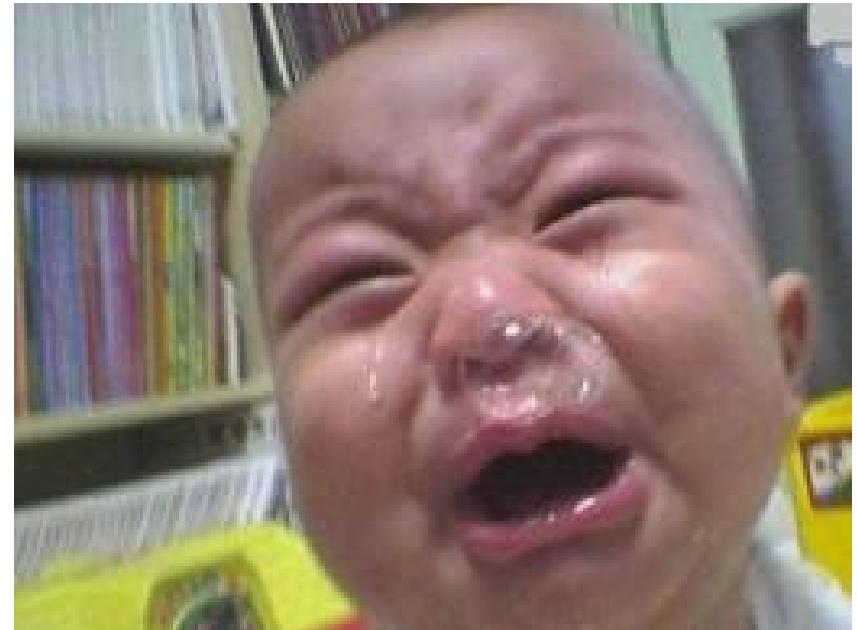


Metagenomics

- Introduction
- Metagenomics
 - Origins
 - Methodologies
 - Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Modern Methodologies

- 1) Select an ecosystem
- 2) Extract DNA
- 3) Sequence
- 4) Cry (Non-optional)



- Introduction
- Metagenomics
 - Origins
 - Methodologies
 - Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

- 1) Select an ecosystem
- 2) Extract DNA
- 3) Sequence
- 4) Cry (Non-optional)
- 5) Analyze

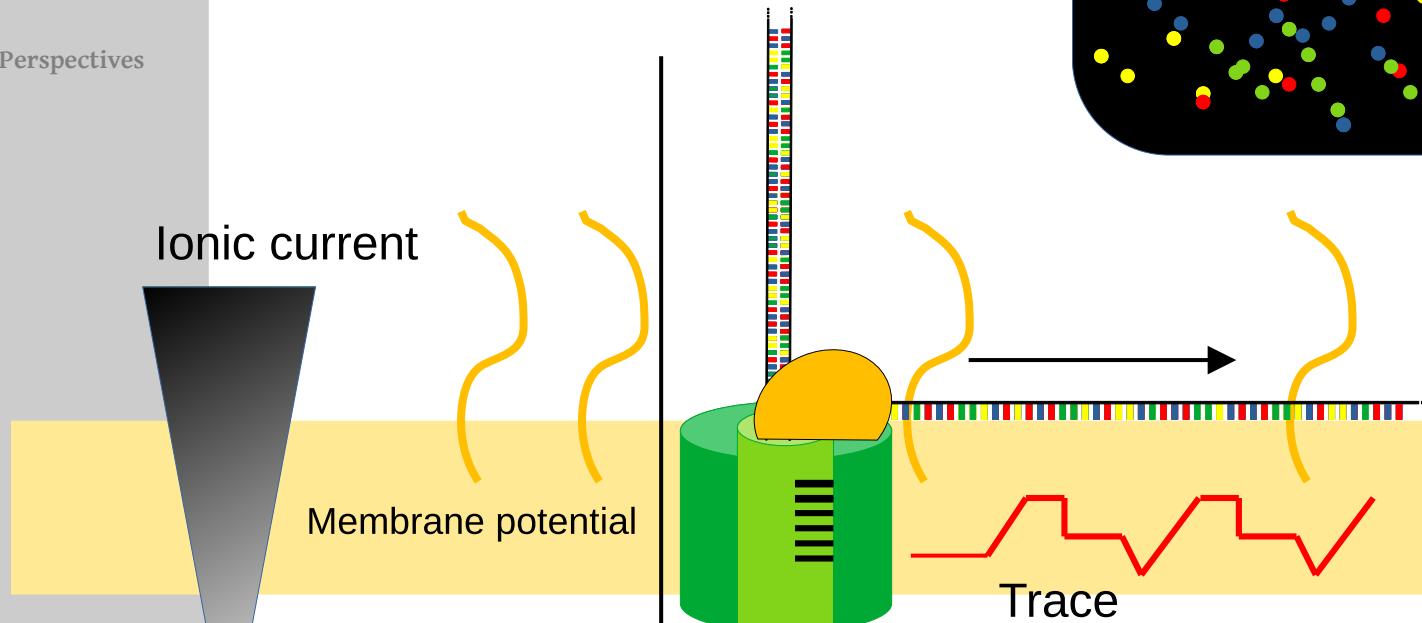
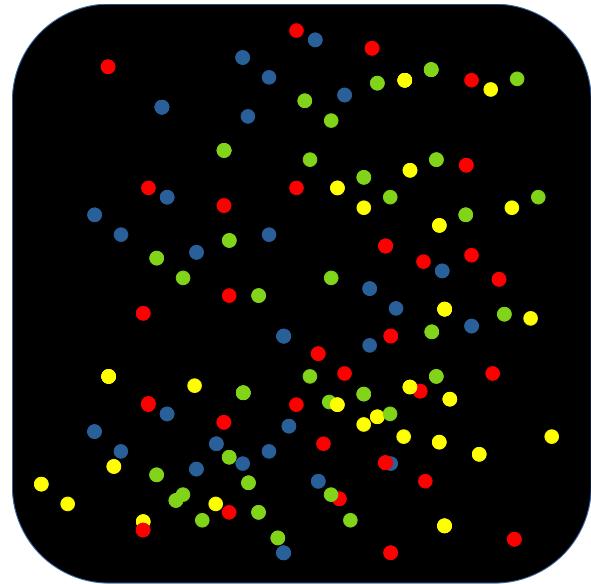


SAGAGUUUGAUUAUGGCUCAGAUUGAACGCUGGCGGAGGCCAACACAGCAAGUOLGAGCGGUACAGGGGGAGCUGG
UCUCUGCUGACGGAGCGCGGAGCGGUACACGUAGGAUCUGCCUGGUAGUGGGGAGAAUCGCCGGAGAAAUCCGGAA
UUAAUACCGCAUACGCCAAGGGGGAAAGCAGGGGAUCCGUUCUUCGGAGAGUGGACCUUGCACAUJGGAUGAGCC
UGCGUCGGAUUAGCUUJGUUGGUUGGGGUAAAGGCCUACCAAGGCAACGAUCGGUAGCUGGUUCAGAGGAUGAUCAGCCAC
ACUGGGACUGAGACACGGCCCAGACUCUACGGGAGGCAGCAGUGGGAAUUAUGCACAAUGGGGAAACCCUGAUGCA
CCAUGCCGCGUGUGUGAAGAAGGCUCUAGGGUUGUAAAGCACUUUCAGCGAGGAGGAAGGUUGUJGAUAAAUCUCAA
AGCUGUGACGUUACUCGCAGAAGAAGCACCAGGUACUCGGUAGCAGCCGCGGUAAUACGGAGGGUGCGAGCGUUA
UCGGAUUUACUGGGGUAAAGCGUGGUAGGGGGCUGGUAAAGUUGGAGUGAAAGCCCGGGGUUACCCUGGGAACUGG
AUCCAAAACUGGGUAGCUAGAGUGCGGAAGAGGAGUGUGGGAAUUCUGUAGCUGGUGAAAGCGGUGAAAGCGUAGAU
AACACCAAGUGGCGAAGGCAGACACUCUGGUACUGACAGCAGGUACAGCAGGGGAGCAACAGGAAGUUAAGAU
CCCUGGUAGGUCCACGCCGUAAACGAUGGUACUAGUUGUCCGGAGGUUUGUAAUCUGGUAAACGAAGCUACCGCAU
UAGACCGCCUGGGGGAGUACGGCCGAAGGUAAAACUAAAUGAAUUGACGGGGGGCCGCACAAGCGGUGGGAGCAU
UUGGUUUAAUUCGAAGCAACGCGAAGAACCUUACUGGUACUCCUGCGAAUCCUUUAGAGAUAGAGGAGUGCCU
CGGGAACGCGAGAGACAGGUUGCGUAGGGUUGGUACUGGUAGCAGCUGGUAGUAGUUGGUUAAGUCCCGCAACGAGCG
ACCCUUGGUCCACGUUACAGCACGUUAUGGUUGGGACUCUGGGAGACUGCCGGUACAACCGGAGGAAGGUUGGGGA
GACGUCAAGUCAUCAUGGCCUUACGGCCAGGGGUACACAGGUGGUACAGACGGUUGCCAAAGACGGUUGCCAAAGCG
GGUGGAGCUAAUCUGAGAAAGUGCAUCGUAGUGGUACUGGUAGUUGGAGUCUGCAACUCGACUCCAGAAGUGGGAA
AUUCGUAAUCGAAUUCGACCGGUACUCCUGGUAGUUGGAGUCUGCAACUCGACUCCAGAAGUGGGAAUCGCUAGU

Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Shotgun Sequencing



- Introduction
- Metagenomics
 - Origins
 - Methodologies
 - Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Bioinformatic analysis

- What is your question ?
- What are you looking for ?
- What are your hypotheses ?
- How will test them ?

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Bioinformatic analysis

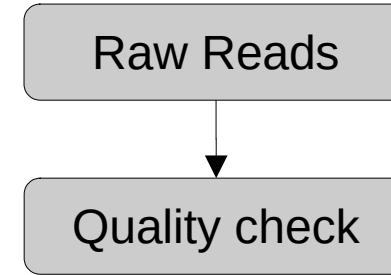
Stereotypical workflow

Raw Reads

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Bioinformatic analysis

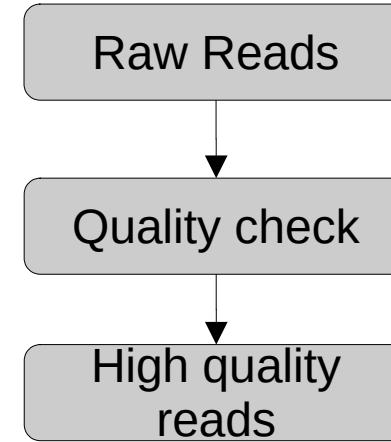
Stereotypical workflow



- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Bioinformatic analysis

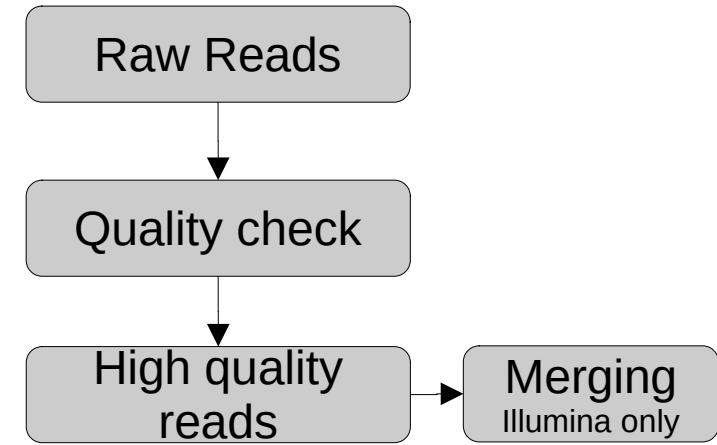
Stereotypical workflow



- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Bioinformatic analysis

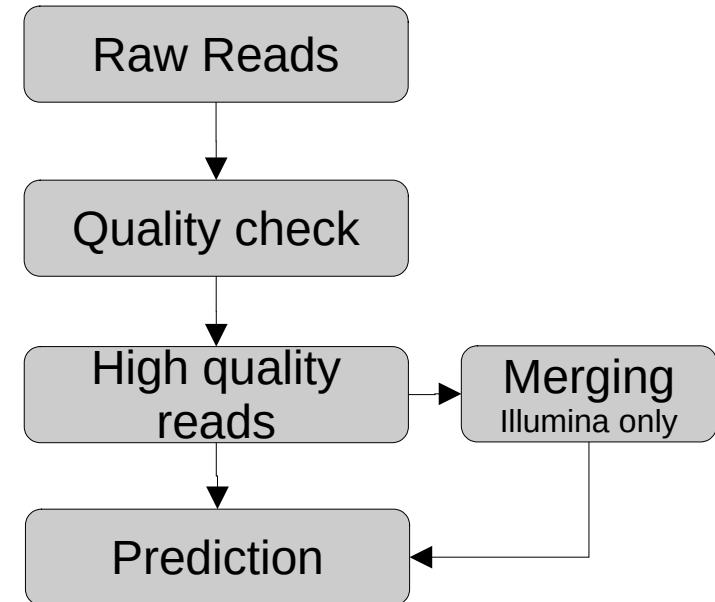
Stereotypical workflow



- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Bioinformatic analysis

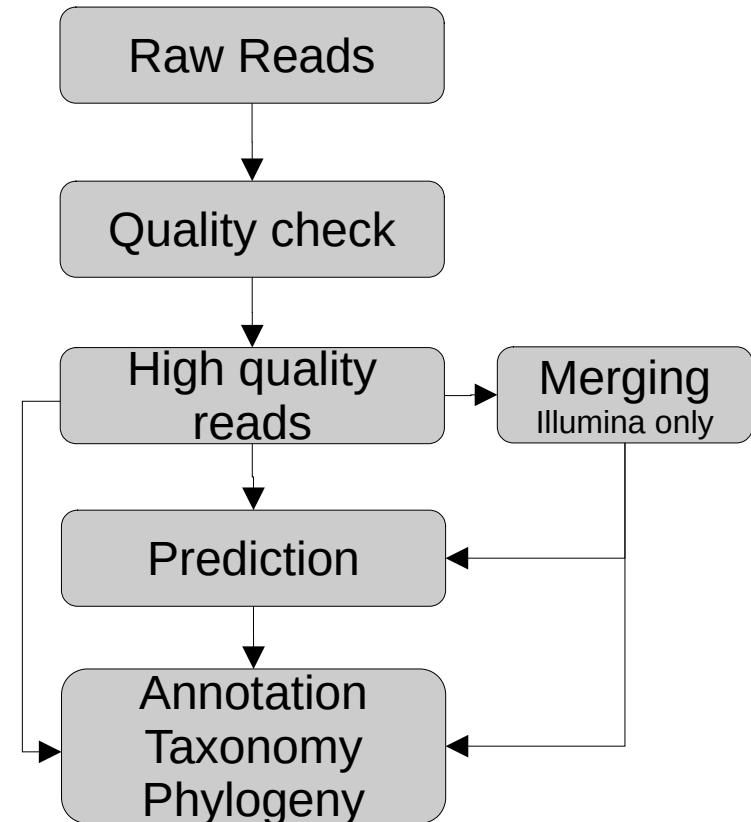
Stereotypical workflow



- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Bioinformatic analysis

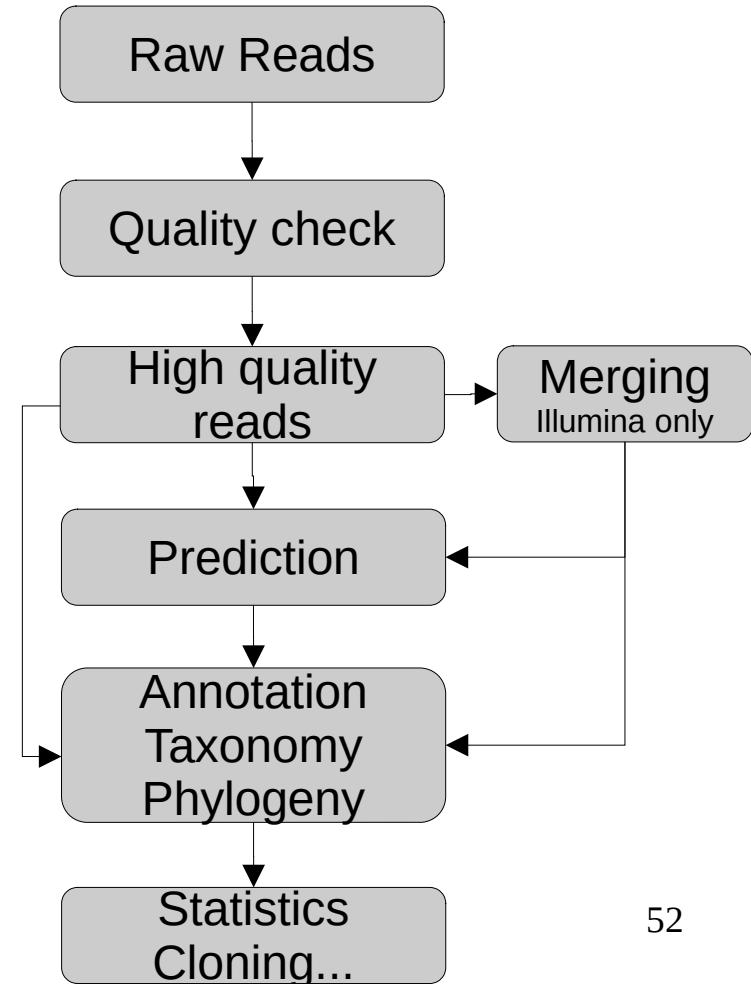
Stereotypical workflow



- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Bioinformatic analysis

Stereotypical workflow

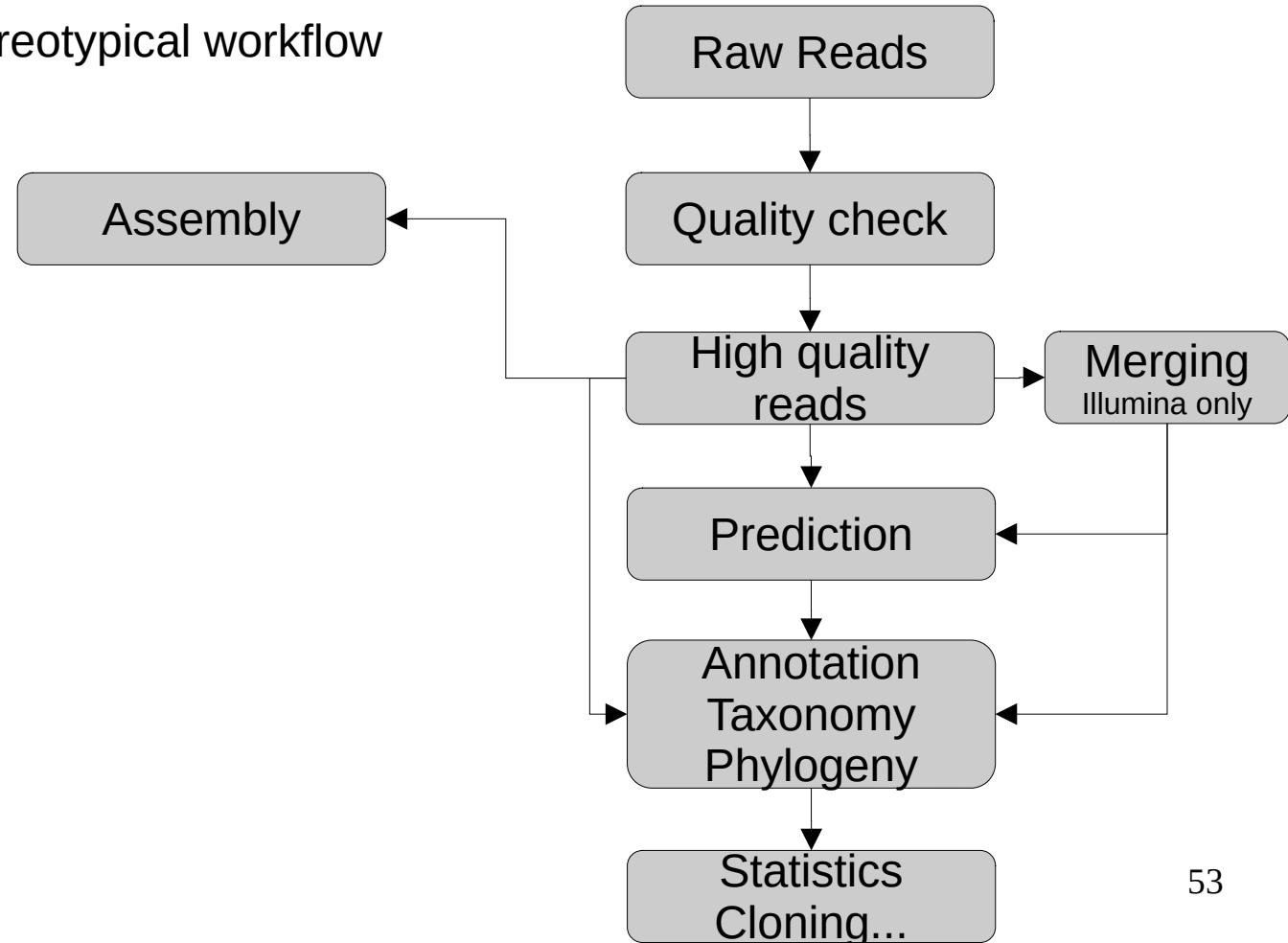


Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Bioinformatic analysis

Stereotypical workflow

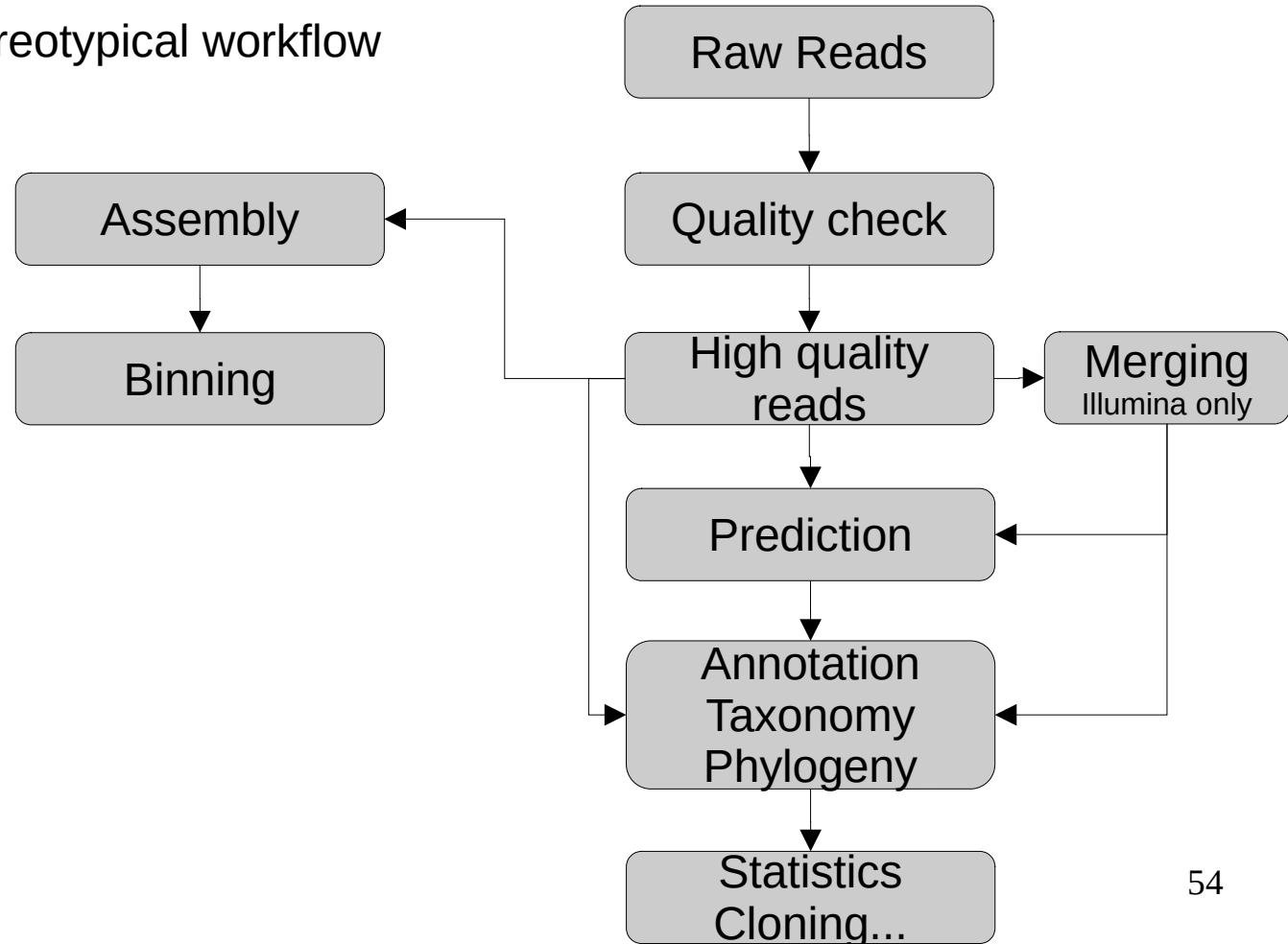


Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Bioinformatic analysis

Stereotypical workflow

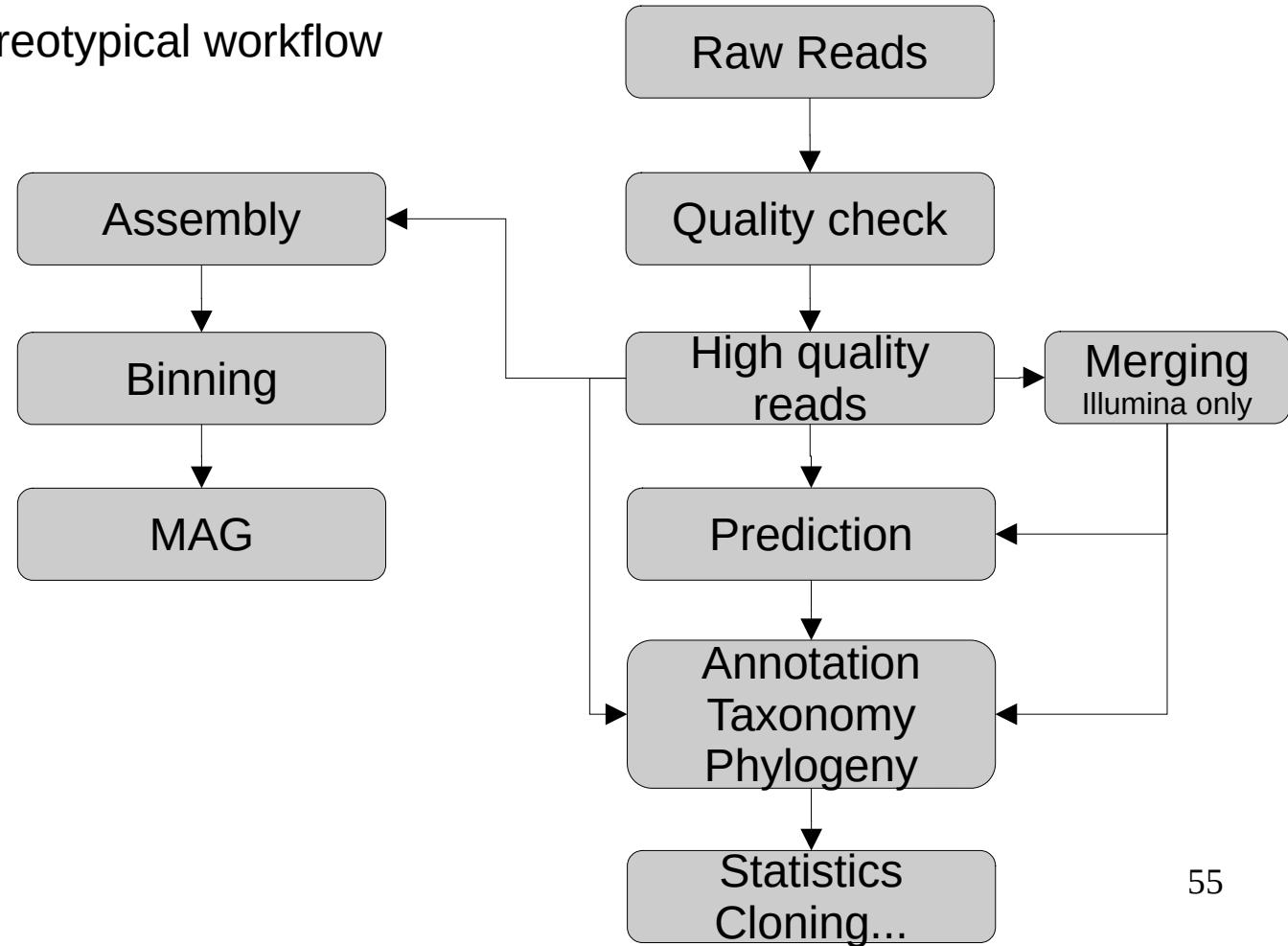


Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Bioinformatic analysis

Stereotypical workflow

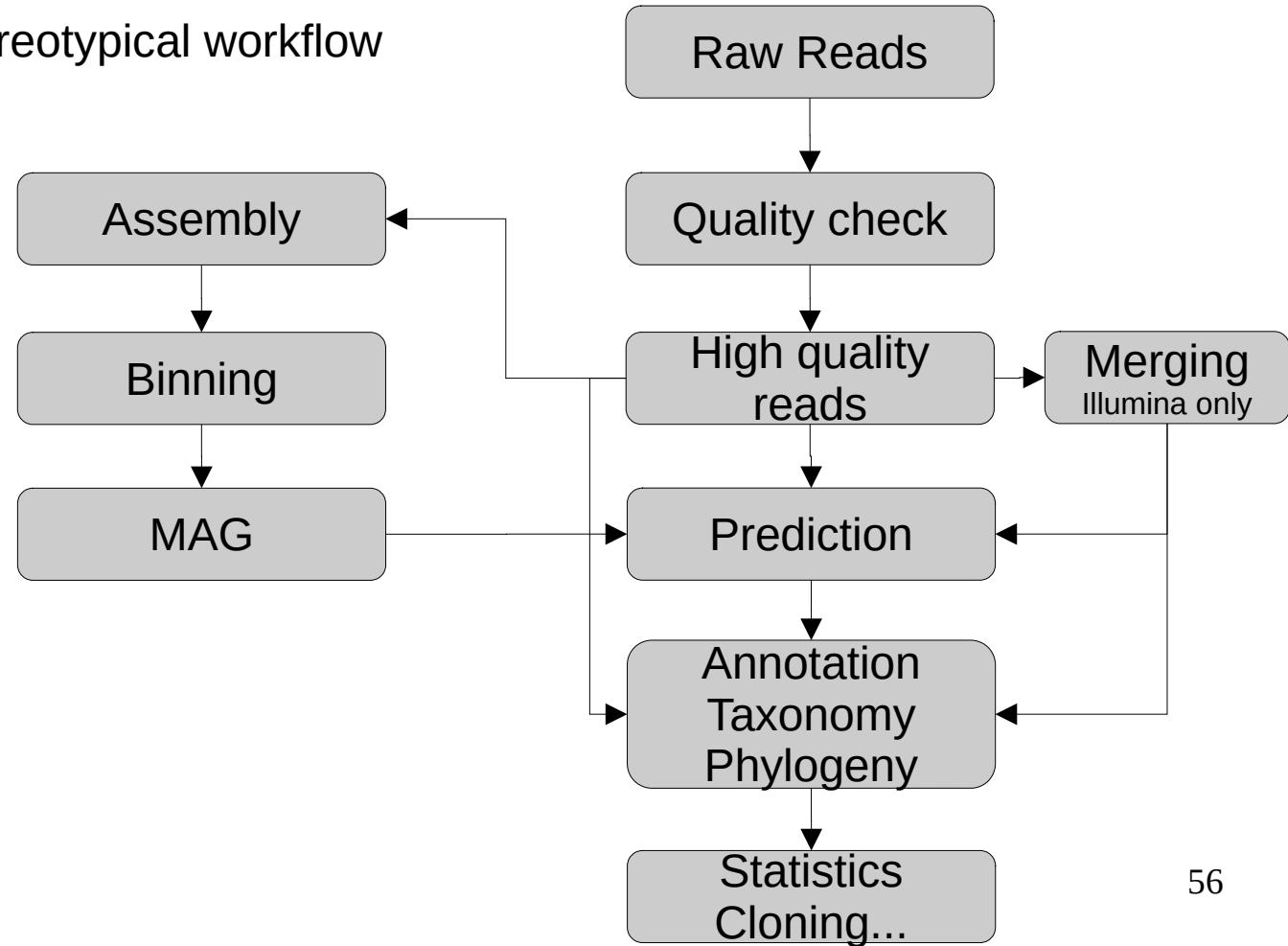


Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Bioinformatic analysis

Stereotypical workflow

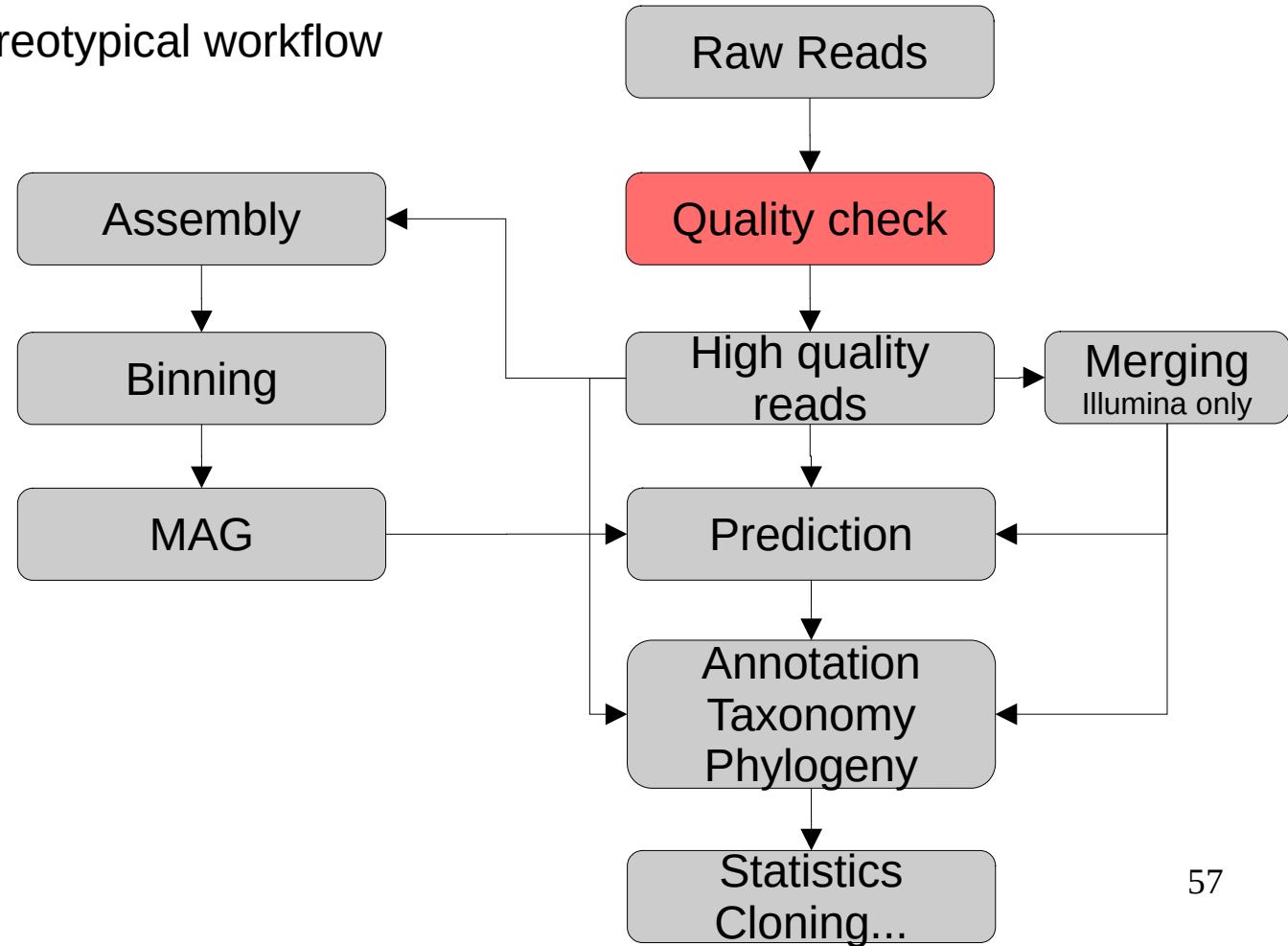


Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Bioinformatic analysis

Stereotypical workflow



- Introduction
- **Metagenomics**
 - Origins
 - Methodologies
 - Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

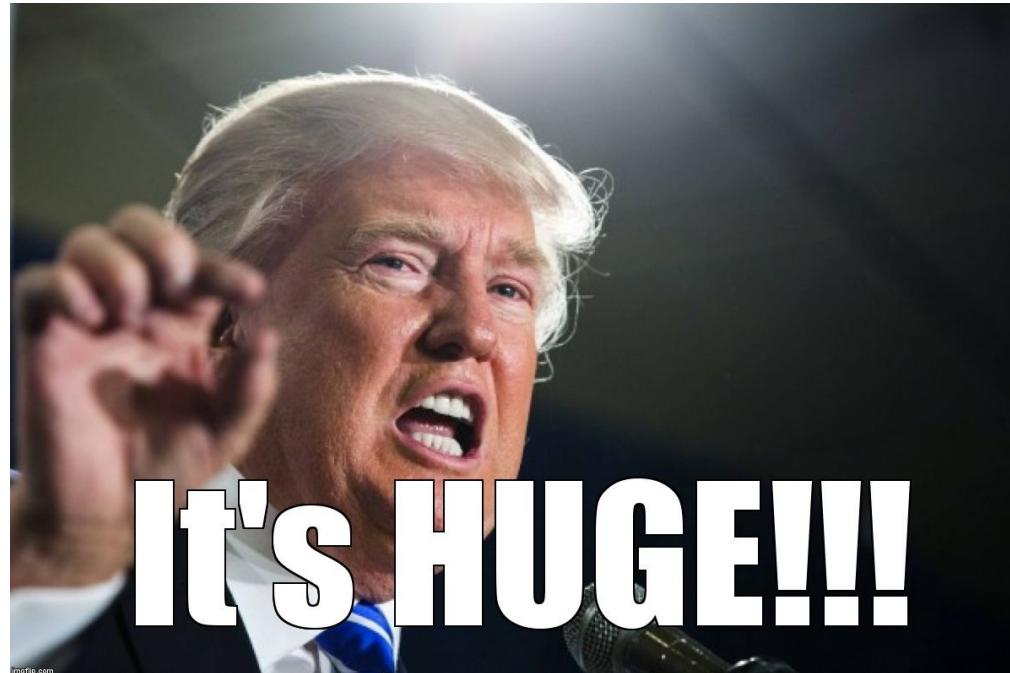
How big is a metagenome ?

Metagenomics

- Introduction
- Metagenomics
 - Origins
 - Methodologies
 - Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Bioinformatic analysis

How big is a metagenome ?



- Introduction
- Metagenomics
 - Origins
 - Methodologies
 - Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Bioinformatic analysis

My last metagenome (1 sample only) is 35 GigaByte large
Around **9.8×10^6 sequences** produced for **32.5×10^9 base pairs**

A4 page, Arial size = 12pts :
~2800 symbols per page
0.1mm thickness

So 11607143 pages
5803572 recto-verso sheet

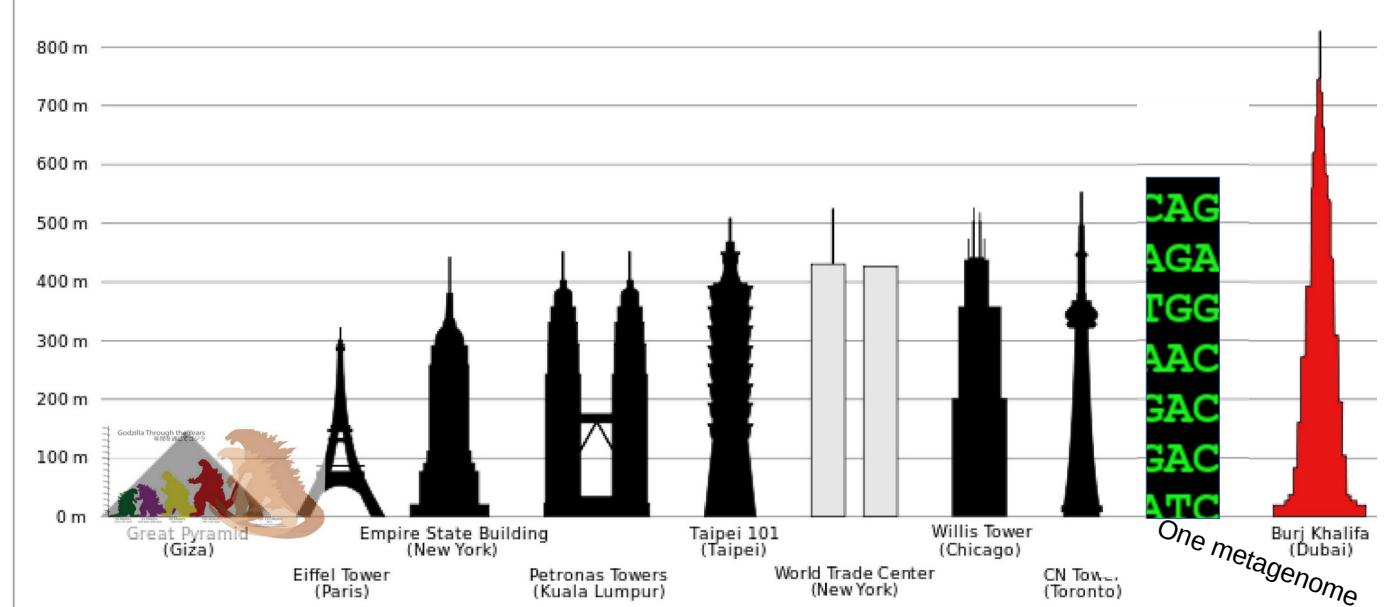
580meter high paper tower

$36m^3$ (43.2 tons of paper)

Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Bioinformatic analysis



I – Metagenomics : Fastq files and quality check

- Introduction
- **Metagenomics**
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

Difference between fasta and fastq ?



Metagenomics

- Introduction
- Metagenomics
 - Origins
 - Methodologies
 - Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Fastq files and quality check

```
032a1dd49-3cfc-47da-bb7f-bc983d9ca2d8 runid=85124c13aa617de74894f30b1c1582ec5f3c1221 read=6596 ch=292 start_time=2023-09-08T21:32:17.473  
399+01:00 flow_cell_id=FAX03081 protocol_group_id=second_comp_mada sample_id=sec_comp_mada parent_read_id=32a1dd49-3cfc-47da-bb7f-bc983d  
9ca2d8 basecall_model_version_id=2021-05-17_dna_r9.4.1_minion_96_29d8704b  
CCTTGACTTCGGTCAAGTACGTTACTCAGGCCGTGACGTGAGCCGGCCGGCGTCCCGTGGCGAACAGCTGGCACCCACACCCACTACGGGGATCTGGGGATCGAGACGCCGGAAACATCGGCC  
TGATCGTCACTAGCGCCAGCTACTCAGGTCAACCCCCGGCTTCATCGAGCGCGTATCGAGCCGGTAAAGGGATGGCAAGCTCGAGCTACATCGGGAGACGGTCAAGTACGCGGCAC  
TGAGCGAGGAATAGGCCGACACGCCCTTGCGCTTGGTAACGTTTTTGCGCTTGCGCTGGCAAGACGCTTGTGATGACGCCGAAGGGGAGTCACCTGATGGCGTTGTCATCACAGCTGCGTGA  
CCCTTCTGAGCACAGCACGCCAGGCCAGGGCCGGCTGCGCTGATGGATGCAACATCGGGAGACGCTGAGCCGGTGCAGACGGCTGCGCTGGCAAGGGGAGTCAGCGCAC  
AGTGTGTCAGTACGGTACGGCAAGGTCACTACGTTGAGCCGGCAACATCGAGCTACGGCTGACGGGACTCCCGCCCTGGCATGTCAGGGGAGACCTTCACTGCGCTTCAC  
GCGGGCAGCAACAGGAGCACCAACCTCGACCCGCAACCCCATCGTAGCTGGCCGACGACTGAGCTGGCCGAGGTGATCGCGTCAAGGGCCCTGCGCTGGCGAGACATCGCCA  
TCGCGGATCATGCTTCAGCGCTTCAGCTTGAAGACGCCATCTGCACTCCGAAGGGCTGTCGCGCAAGACTCCCGCTCCATGCAACAAAAGATGAACTGAGGGCGTGA  
TCACCTGATACTCCGGGCCTCGAGTGAGGGCCGCTGCCCTGACCTGAGGGGAGACGGCATCTGCGTCAAGGGGAAATCAGCCGACATCTGGTCCGAGGCTGGCGAG  
GAGGGCCTTCGCGTTGATCTGGTGAGGGCCGGAAGTGAAGGGACACCTGCGTCCGGTCCGGTCCGGTCCGGTCCGGTCCGGTCCGGTCAAGGGCGAGGCCGACCCCCGAA  
CCCAGCGCTCCGAGATGGTGGCGGTGACGTGGGCCAAGAGCTGCAAAGCTGCAGGTGGCGACAAGGGCCAACCGGCCAGGGCGTGGCTCCAAGAATCATGGCCCAGGACA  
GAAGACGGCACCCCGTCGATCTGGTCTTAACCCCGTGGCGTGCCTCGGCATGAACCTGGCGGAATCGAAACCCACTGGTGGAGGTGCCCCACCTGACCCGAGAGCAATAC  
+  
(%$((()34@=<:;%%1=<6940.-.---4,+0/7,,3//,7//,99,37790'!!!'.(7.>?A95+&'22(+710111,/+,-2325*(,*%&'3())/01-,,,)(-3087()+.+,.4201++5+67:43+'  
(&-.57,-0+''')*2..1//0:;&67929<<-899692332)2>5461=99878?=>?:=>?:=>?:&09577740.01311&&2000+)+''245..0.11237965556:90$%(.2.567863+)&8645..  
/.31(&&&*&'+*,&$#&$#&$#&($#&$#&$#&%*'0232-+)(%+&(*-$#&$#&$#&%*'*$*,*)(&&.,+.,,)(&&.*+078)6((+  
3222;:'&&567';';=:(94)))'55<=:;:=89:7744320555'(4)63:+*+'&767';=987:4304(+)&&666,*+2;;:<G;??;98:>*8:./068>9??<>923599855259430  
)0/05368=<++18==AD>0,-*/-.55575;A>:964:9999:223<,:=((*)//036469=87566:214444,8=99=-1:)))*8(+2487)-(8665&&&)2554+%%'*05;1778:66  
&A7B@>322,%.59)3//005311-897)(157DBB2=-/7/K/-112;.;A@=945328...5/160/A@87;(''-7979=-;<C(C;:=101A@?>ABAE?@?<78/3+**+,16.5001)()8  
87(**(*.,67,,/98431+),(.,/5@0/>687,(.'1136:991'+*7899;978:957.0211192(+(-))'*&%+0--89),9,222;><77:5557:6..1656.1,-.1.57785588<  
856446559-/7819AB:<:C>11;&'+065--+,&&)+4:85=/974323//...7:878&*&92348:&2;</--/01--44:22/&&*2,,28:??<:76321225'/'(,55(*(.  
,,1343587:/.32711131.08:667/0041,-'')..-/34*!!),-,0:21761577858:=<878=-.577620-6*)(5)((156=>:3567445,)(&...,*+541./01221'8<0  
87BA:9999;,<=84'(&*&23499;9;38994,*126550,322336:1254,++())*459,+/,7(''/(057879:<862/-2:=6840,.129;<,6,-'&/'1572,
```

Metagenomics

- Introduction
- Metagenomics
 - Origins
 - Methodologies
 - Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Fastq files and quality check

Header

```
| 832a1dd49-3cfc-47da-bb7f-bc983d9ca2d8 runid=85124c13aa617de74894f30b1c1582ec5f3c1221 read=6596 ch=292 start_time=2023-09-08T21:32:17.473  
399+01:00 flow_cell_id=FAX03081 protocol_group_id=second_comp_mada sample_id=sec_comp_mada parent_read_id=32a1dd49-3cfc-47da-bb7f-bc983d  
9ca2d8 basecall_model_version_id=2021-05_17_dna_r9.4.1_minion_96_29d8704b  
CCTTGACTTCGGTCAAGTACGTTACTCAGGCCGTGAGCTGAGCCGGCCGGCGTCCCGTGCACGCGTGCAGCTGGCCATCTGGCCGATCTGGCCGATCTGGCCGAGAGACGCCGGAAACATCGGCC  
TGAGCTCATCAGCGCCAGCTACTCAGGTCAACCCCTGGCCTTCATCGAGCCGGTATCGAGCCGGTGAAGGATGGCAAGGGTCAAGCTACGCGGAGACGGTCACTACGCGGAGACATCGGCC  
TGAGCAGGAATAAGCCGGCACACAGCTTCTGGCCTTGCTAACGGTTTCTGGCTTGCAAGACGCTTGTGATGACGCCGAAGGGGAGTCACCTGACCTGATGGCCTTGTCCATCACAGCTGCGTGA  
CCCTTCTGGAGCACAGCACGCCAGCCAGGGCCAGCTGCGCTATGGATGCAACATGGGCTGAGACGCTGACGGCTGCGCCTGTCGGCTAACAGGCCGGCGTGGCAAGGGTCAAGCCACTCGGCC  
AGTGTGCTAGTACGGTACGGCAAGGTCACTACGTGGACCCGCAACATCGAGCTACGGCTGACGGGACTCCCTGGCCCTGGCATGTGCAAGGGGAGACTTCACTGGCTTCACTCGGCC  
GCGGGCAGCAACAGGAGCACCAACCTCAGCCGGCACCCCCATCGAAGCTGGGGAGACGACTGCACTGGCAGGGAGCTGCACTGGGCTGACGGCCCTGCACTCCGACCTGGGCTGGCAGAACATCGGCC  
TCGCGGATCATGGCTTCAGCTTGAAAGACGCCATCTGCACTCCGAGGGCTGTGCGCAGGACTTCCCCTGGCATCTGAAAAGATGAACTGAGGGCTGATCAAAGCAGGGCCGGAA  
TCACCTGATATCAGCTGGGCTCTGGATGAGGGCCGGCTGGCTGACCTGGGAGACGAGCTGGCAGGCTGGGAGACGGCATCTGGGCAACGGCTGACGGGCTGGCTCAAGGGCAGGCC  
GAGGGCCTCTGGTTCAGCTGGTGAAGGGCCGGAGGGACACCTCGCTGGGGCTGGCTGGGGGGGGCATCTGGTGTGCAAACCTGGCTGGGCTGGGCAATCGAAACCCACCTGGG  
CCCAGCGCTCCGAGATGGTGGCGGTGACGTGGGCCAAGAGCTGCAAAGCTGCAAGGTGGGCAACGGGACAGGGCTGGCTCCAAGAATCATGGCCGAGGACAATGGCTACCTG  
GAAGACGGCACCCCGTCACTGGTGTGCAACCCCTGGGCTGGGCAATCGAAACCCACCTGGGCAATCGAAACCCACCTGGGCTGGGCAATCGAAACCCACCTGGG  
+  
(%$((()34@=<::;%%1=<6940-.--4,+0/7,,3//,:/7;;99;37790''''((.7>?A95+&'22(+710111,/+,-2325*,*%&'3())/01-,,)(-3087()+.+,4201++5+67:43+'  
(&-.57,-0+''')*2..1//0:;&67929<<-899692332)2>5461=9987?=>?:=>?:=>?:=>?:=>?:=>?:=>?:=>?:=>?:=>?:=>?:=>?:=>?:=>?:=>?:=>  
/.,.31(&##&#'*+, '$#*$##$#$(($##$##$#%&%'0232-+))+(+$%&-$#%$#%$#%&%'**$%,*) (6&.,+,,,) ((6+*) ('%$##$#%&%')) (%&(-.-*#078) 6((+-  
3222;:'%&#567';,:=(#;94))) ))'56<=:;:==89;7744320555'(4) 63:+*';&767';=987;4304(+) ;&666+,*)+2;:;<:G;??:98:>8*: /068>9??:>923599855259430  
) /05368=<++18==AD>0,-*/-5575;A>:964:9999:223<:==(*0//036469=87566:214444,8=99=-1:)))*8(+2487)-(86656&&)&2554*'%%'*05;1778:66  
&A7B@>322,,%.59)3//005311-897) (157DBB2=:/7/K/-112//;A?945328...5/160/A@87;(''-7979=-;<0C();=101A@?>ABAE?@?<78/3+**+,16:5001() 8  
87((**(*,.67,,/98431+),/,,/5@0/>687,(''1136:991'+7899;978;957.0211192(+(-))'&%*+0--89),9,222;:>77;5557:6..1656.1.-.1.57785588<  
856446559-/7819AB:<:C>11/+'065--+,&%)/4:85/=974323//...7:878&*&92348*'62;</--/01--44://22/&*2,,?8:??<:6321225'/'(,55(*(.  
,,11343587:/.32711131.08:667/0041,-'--..-/34*!'),-,0-21761577858=<878=-.577620-.6*)(1156=>;:3567445,) (+...,*&541./01221'8<0  
87BA:9999;:<84'(&#823499;:38994,* 126550,322336:,1254,++*)*459,+/7(''057879:<862/-2:=6840,.128<:6,-'1'/*1572,
```

Metagenomics

- Introduction
- Metagenomics
 - Origins
 - Methodologies
 - Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Fastq files and quality check

Header

```
@S2aidd49-3cfc-47da-bb7f-bc983d9ca2d8 runid=85124c13aa617de74894f30b1c1582ec5f3c1221 read=6596 ch=292 start_time=2023-09-08T21:32:17.473  
399+01:00 flow_cell_id=FAX03081 protocol_group_id=second_comp_mada sample_id=sec_comp_mada parent_read_id=32aidd49-3cfc-47da-bb7f-bc983d  
9ca2d8 basecall_model_version_id=2021-17_dna_r9.4.1_minion_96_29d8704b  
CCTTGTACTTCGTCAGTTACGTACCTCAGGGCGCTGAGCTGAGCCGCCGCGCTCGCGTGCAGCACGCGTGCAGCTGCAGGCCATCTGGCCGATCAGAGACGCCGGAAACATCGGCC  
TGATCTCATCAGCGCCAGCTACTCAGGTCAACCCCCGGCTTCATCGAGGCCGTATCGGCCGCTGAAGGGATGGCAAGCTGAGCGTCCAGTACGAGGATCGCAAGGTACGGCAGAC  
TGAGCGAGGAATAGGCCGCACACAGCTTCGSCCTGGTAACTGTTTTTCGGCTTGCGCTTGCAAGACGCCCTGTGATGCACGCCGAAGGGGAGTCACCTGATGGCCTTGTCATCAACAGCTGCGTGA  
CCCTTCTGAGCACAGCACGCCAGGCCAGCGCCAGCGCTGCGCTATGGATCGAACATGGCTGAGCAGCTGAGCGTCCCTGCGCTGAACAGGCCGCCGCTGGCAGCCTGGTGCAGCGAC  
AGTGTGCTGAGCTGAGCGTACGGCAAGGTCACTACCTGAGCGCCGCACATCGAGCTACGGCTGACCGGAGCTCCCTGCCCGTGGCATGTCAGGGGAGACCTTCACTGCC  
GCGGGCAGCAACAGGAGACCAACCTCAGCCGCAACCCCATCGTAAGCTGGCCGACGACTGAGCTGGCCGAGGTGACATCGGCTGAGCGCCCTGCACTCCGACATCGGCA  
TCGCGGTATCATGCTTCTGAGGCGTTCACTTGTGAAAGACGCCATCTGCATCTCCGAGGGCTGTGCGCAAGACTTCCCGTCCCTGCACATCGAAAAGATGAACTGAGGGCGT  
TCACCTGTATCTCCGGGCCCTGAGTGAGGCCGCCCTGACCTGAGCGACGGCATCTGGCCGCTGGCCGAAATCAGCGGACATCTGGTCCGAGCTGGCCGAGACATCGGCA  
GAGGGCCTCTGGTTCATGGTGAAGGGCAGGGAACCTCGCTGCGCTGGCCGCTGGGGGGCGCATGTGGTCAAAAGCTGGCCCTTCCGCCGGCGAGCAAGGGCGTGGCT  
CCCAGGCGCTCCGAGATGGTGCGCTGTACGTGGCCCAAGAAGCTGCAAAGCTGCGAGGTGGCGACAAGGGCCAACCGGCCAGGGAACAGGGCGTGGCTCCAAGAATCATGGCC  
GAAGACGGCACCCCGTCACTGGTTCAACCCCGTGGCGTGCCTCGGCATGAACCTGGCCGAATCGAAACCCACCTGGGTGAGGTGCCCCACCTGACCCGAGAGCAATA  
+  
(%$((())34@=<::;%%1=<6940-.--4,+0/7,,3//,7//,99,37790+!!!'.(7.>?A95+&'22(+710111,/+,-2325*(,*%&'3())/01-,,,-)(-3087()+.+,4201++5+67:43+'  
(&-.57,-0+'''(*2..1./0:'&67929<<-899692332)2>5461=99878?=>?:=>?:=>09577740.01311&&2000+)+'!245..0.11237965556:90$%(.2.567863+)_8645..  
/.31(&##'*+,'$#$$##$(#&##'*'$#%'10232-+)(%+$-&$#%$#%$#%$#%'**'$#,*)(&#+,+,,,)((#+)('%%$#&%$'))(%&(-.-*#078)6((+-  
3222;:'%&E567;:;:=#(94)))'!56<=:;:=#89:7744320555'6)63:+*+'!&767';=987:4304(+)&666+,*+2;:;:<G;??;98:>8*: /068>9??<>923599855259430  
)0/05368=<++18==AD>0-*/-.5575;A>:964:9999:223<;==(*0./036469=87566:214444,8=99=-1:)))*8(+2487)-(8665&&2554+'%%'*05;1778:66  
;A7B@>322..%,59)3//005311-897)((157DBB2=:/7/K/-112;:;A?=945328..5/160/A@87;(''-7979=:-<EC();=101A@?>ABAE?@<78/3+**,.16'5001()8  
7;5547991212&(+/4)))03@<-'!654/(''35550013557910//76028:8777-):8652)((899566:446=<81742865:8309,32<,:=764:55..&677<05)-8  
87((**(*..67.,/98431+)*,+/5@0/>687.(''1136:991'+7899;978:957.0211192(+(-))'%%*+0--89),9,222:><77:5557:6..1656.1.-.1.57785588<  
856446559-/7819AB:<:C>11:+'065-+-&%)+4:85=/974323//...7:878&*92348*'62:</--01-+4:22:6&*2,,?8:??<:6321225'!/(',,55*(<..  
,,1343587:/:32711131.08:667/0041,-'')..-/34*!!'),-,0:21761577858=<878-.-.577620-.6*)(156=>:,3567445,)(+-,,,*+541./01221'8<0  
87BA:9999;:<84'(&#823499:9;38994,*126550,322236:1254,++()&459,+/,7(''057879:<862/-2:6840.,129<,6,-'f'/*1572,
```



Metagenomics

- Introduction
- Metagenomics
 - Origins
 - Methodologies
 - Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Fastq files and quality check

Header

```
@S2aidd49-3cfc-47da-bb7f-bc983d9ca2d8 runid=85124c13aa617de74894f30b1c1582ec5f3c1221 read=6596 ch=292 start_time=2023-09-08T21:32:17.473  
399+01:00 flow_cell_id=FAX03081 protocol_group_id=second_comp_mada sample_id=sec_comp_mada parent_read_id=32aidd49-3cfc-47da-bb7f-bc983d  
9ca2d8 basecall_model_version_id=2021-05-17_dna_r9.4.1_minion_96_29d8704b  
CCTTGACTTCGGTCAAGTACGTTACTCAGGCCGTGACGTGAGCCGGCGCCGGCTCGCGTGCACGCGTGAACGGTGCACCCACTACGGGGATCGAGACGCCGGAAACATCGGCC  
TGATCCTACAGCGCCAGCTACTCAGTCACCCCCGGCTTCATCGAGCGCGTATCGCGCGTGAAGGATGGCAAGGTCACGGTCACTACGCGGACATCGGGATCGAGGATCGTAC  
TGAGCAGGAAATAGCCGCACACGGCTTCGCGCTGGTAACTGTTTTGCGCTTGCGCTGCCAAGAACGCGTTGTGATGCCACGCCGAAGGGGAGTCACCTGATGGCGTTGCCATCACAGCTGCGTGA  
CCCTTCTGAGCACAGCACGCCAGCCAGCGCCAGCCTGCGCTATGGATGCAACATGGCGTGCAGACCGCTGCCCTGCGCTGCCAGACGCCGGCTGGCAAGGGTGTGACGCCGGCGTGTCA  
AGTGTGCTAGTACGGTACGGCAAGGTCACTACGTGGACCCGCAACATCGAGCTACGGTACGGGACTCCCGCCCTGCCATGTGCAAGGGGAGACTCCACTGCCCTGCGCTTAC  
GCCGGCAGCAAACAGGAGACCAACCTCAGCCGACCCCCATCGAAGCTGGGGACGACTGAGCTGCCGAGGTCATCGCTGACGGCCCTGCACTCCGACCTGGCCAGAACATCGCCA  
TCGGGATCATGCCCTCGAGCGCTCAGCTTGTGAAAGCGCATCGCATCGCAGGGCTGTGCCAGGGACTCCCGCTCCATCGAAAAGATGATACTGGAGGGCGTATCACCAACGGGGCGGA  
TCACCTGATATCCGGGCGCTCGAGTGAGGGCCGCTGCCCTGACCTCGAGGACGACGGCATCTGGCCGCTGCCGGGAAATCACGGCGACATCTGGTCAGGGCCCTGCGCTGGCCAGAACATCGGA  
GAGGGCCCTCGCGTTCGATTCGAGGCGGAAGGACACCTCGCTGCCGTGCCGGCGCATGTGGCTAACACGGTGGCGCTTCCGCCGCGCGACAGGGCGTGGCTCCAAGAATCATGGCCCAGGACAATCG  
CCCGCGCTCCGAGATGGTGGCGTGTACGTGGCCAAGAGCTGCAAGCTGCGAGGTGGCGACAAGGGCCAACGGGACAGGGCGTGGCTCCAAGAATCATGGCCCAGGACAATCG  
GAAGACGGACCCCGTCACTGGTGTCAACCCCGTGGCGTGCCTCGCGCATGAACTGGCGGAATCGAAACCCACCTGGGTGAGGTGGCCACCTGACCCGAGAGCAATA  
+  
(%$((()34@=<::;%%1=<6940-.--4,+0/7,,3//,:/7;;99;377930!!!!!(.7>?@A95+&'22! (+710111,/+, -2325*, *%&'3 ()/01-, ,) (-3087 ()+.+4201+-+5+67:43+!  
(&-, -57,-0+''!) *2,.1//:0!:&67929<<-899692332)2>5461=99878?=>?:=0@9577740.01311&&2000+)+'!245..0.11237965556:90$%(.2.567863+))8645..  
/.31(&&&*&!*+, '$#&$##$#($#&$##$%&'0232-+)(+%&(-$#&$#&$#&$%&'**'$#, *)*(6&., +., ,) ((6+*)(%$&$#&$#&$%)) (%&(-.-*+078)6((+  
3222;:'!&&567;:,;:=94:))'!56<=:;:=89:7744320555'6)63:+*+'!&'767'=987:4304(+)&&66+,*+2;:;<:G;??:98:>*:8*/.068*9??<>923599855259430  
)0/05368=<++18==AD>0-,*-/5575;A>:964:9999:223-<:==(*0./036469=87566:214444,8=99=-1:)))*8(+2487)-(8665&&&2554*'%%'*05;1778:66  
;A7B@?>322,.%,59)3//005311-897)((157DBB2=:7/K/-112;,:A?945328..5/160/A@87;(''-7979=:;<@C();=101A@?>ABAEE?@<78/3+**+,16:5001)()8  
77;5547991212&&(+/4)))03@-<!'654/(''35550013557910//76028:8777-):8652)((899566:446=<81742865:8309,32<:=764:55..-66@77<05)-8  
87((**(*.,67.,/98431+)*,/5@0?>687.(''1136:991'+7899;978:957.0211192(+(-))'&&*+0--89),9,222:><77:5557:6..1656.1.-.157785588<  
856446559-/7819AB:<:C>11:&'065--+,&&)/4:85=/974323//...7:878&&'92348':&2;<(/-'-01-+4:22/&&*2,,?8:??<:76321225'/-'(,55(*(.-(  
.1343587:/.32711131.08:667/0041,-'-)..-/34*!!),-,0-21761577858=<878-.-577620-.6*)(156=>:3567445,) (+..., *+541./01221'8<0  
87BA:9999;:=84'(&&23499:9;38994, *, 126550,322236:1254,++)()&459,+/,7(''057879:<862/-2:=6840,.129;<,6,-'f'/*1572,
```

Sequence

Legacy

Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Fastq files and quality check

Header

```
@S2aidd49-3cfc-47da-bb7f-bc983d9ca2d8 runid=85124c13aa617de74894f30b1c1582ec5f3c1221 read=6596 ch=292 start_time=2023-09-08T21:32:17.473  
399+01:00 flow_cell_id=FAX03081 protocol_group_id=second_comp_mada sample_id=sec_comp_mada parent_read_id=32aidd49-3cfc-47da-bb7f-bc983d  
9ca2d8 basecall_model_version_id=2021-05-17_dna_r9.4.1_minion_96_29d8704b  
CCTTGACTTCGGTAGTTACGTACGGCGCTGAGCTGAGCCGGCGGCCGCTCCGCTGGCACGAGCTGACCACACCGACTACGGGGATCTGGGGATCGAGACGCCGGAAACATCGGCC  
TGATCTCATCAGCGCCAGCTACTCAGGTCAACCCCTGGCTTTCATCGAGCCGGTATCGCCGAGTCAGGATGGCTACGGTACAGCTACGGCAC  
TGAGCAGGAATAGCCGCACACAGCTTCTGGCTTGGTAACTGTTTTCTGGCTCTGGCTTGCAAGAACGCTTGTGATGACGCCGAAGGGGAGTCACCTGATGGCTTGTCCATCACAGCTGCTGATT  
CCCTTCTGAGCACAGCACGCCAGCCAGGGCTGGCTGATGGATGCAACATGGCTGAGCACGCTGAGCCGGCTGGCTGAGCCGGCTGGCTGAGCAGCCGGCTGGTCAACGACTCAGGCAC  
AGTGTGCTAGTACGGTACGGCAAGGTCACTACCTGAGCCCCAACATCAGCTACGGTACGGGACTCCCTGGCCCTGGCATGTCAGGGGAGACTTCACTGGCTTAC  
GCGGGCAGCAACAGGAGCACCACCTCGACCCGACCCCCATCGTGAACGTGGGGAGCAGCTGAGCTGGGGAGCTGACCTGGCCCTGACCTGGCCAGAACATCGCCA  
TCGCGGATCATGCTTCTGAGGGCTCAGCTTGTGAAAGACGCCATCIGCATCTCCAGGGCTGTGCGCAAGACTTCCCGTCTCCGACATCGAAAAGATGAACTGAGGGCTGATCAA  
TCACCTGATATCCGGGCGCTCGAGTGAGGGCCGCTGGCTGACCTGAGGGAGCAGCTGGCCGAGGCTGACGGCCGACATCTGGTCCGAGGCTGGCCGAGGGCAAC  
GAGGGCCCTCTGGTTCATGGTGAAGGGCGGAAGTGAACTCGTGGCCGCTGGCTGGGGCGCATGTGGTCAAAACGGTGGCCCTTCCGGCCGGAGCAGGGCTGAATCAAG  
CCC GGCGTCCCGAGATGGTGGCGTGTACGTGGCCAAGAAGCTGCAAAGCTGCAAGGTGGCGACAAGGGCCAACCGGCCAGGGGAGCTGGTCTCAAAGATCATGGCCCAGAGCATGGCTTACCTG  
GAAGACGGCACCCCGTCACTGGTCTTCAACCCCGTGGCGTGGCCATGAACCTGGCGGAATCGAAACCCACCTGGTGAAGGGTGGCCACCTGACCCAGAGCAATAC  
+  
(%$((()34@=<:;%%1=<6940-.--4,+0/7,,3//,:/7;;99,377930!!!!'(.7>?A95+&'22(+710111,/+-2325*,*%&'3())/01-,,)(-3087()+.+,4201++5+67:43+'  
(&-.57,-0+''')*2..1//0:;!&67929<<-899692332)2>5461=99878?=>?:=09577740.01311&&2000+)+''245..0.11237965556:90$%(.2.567863+)&8645..  
/.31(&6&&*&#'*+, '$#*$##$#$(#&#*$##$%'10232-+)+(%&(-$#*$##$#%&*'**'$#,*)*(6&.,+.,,)(6+*)(%$##$#&%))($&(.,-*+078)6((+  
3222;:'!&6567;';:=c(94)))''56<=:;:=89:7744320555'4)63:+*+'&767'=987:4304(+)&666+,*)+2;:::<G;??;98:>8*: /068>9??<>923599855259430  
)/05368=<++18==AD>0,-*/-.55575;A>:964:9999:223<,:==(*0./036469=87566:214444,8=99=-1:)))*8(+2487)-(86656&&2554+''%&*05;1778:66  
&A7B@>322,%.59)*3//005311-897)((157DBB2=-:7/K/-112;/:A?945328..5/160/A@87;(''-7979=:;<0C();=101A@?>ABAE?@?<78/3+**+,16:5001)()8  
77;5547991212&&(+/4)))03@-<-'654/(''%35550013557910//76028:8777-):8652)((899566:446=<81742865:8309,32<:=764:55..&67=877<05)-8  
87((**(*,.67,,/98431+),(./5@0/>687,(''1136:991'+7899;978:957.0211192(+(-))&%*+0--89),9,222:><77:5557:6..1656.1.-.1.57785588<  
856446559-/7819AB:<:C>11/+065--,&&)/4:85=/974323//...7:878&*92348*62;</--/01--44:/22/&&*2,,?8::??<:76321225'/-'(,55(*(.  
,,11343587:/.32711131.08:667/0041,-'')..-/34*!!),-,0:21761577858=<878=-.577620-6*)(5)((156=>:,3567445,) (+...,*+541./01221'8<8  
87BA:9999;:<84'(&823499:9;38994,*126550,322236:1254,++())*459,+/7(''(057879:<862/-2:=6840,.129<,6,-'f'/*1572,
```

Sequence

Legacy

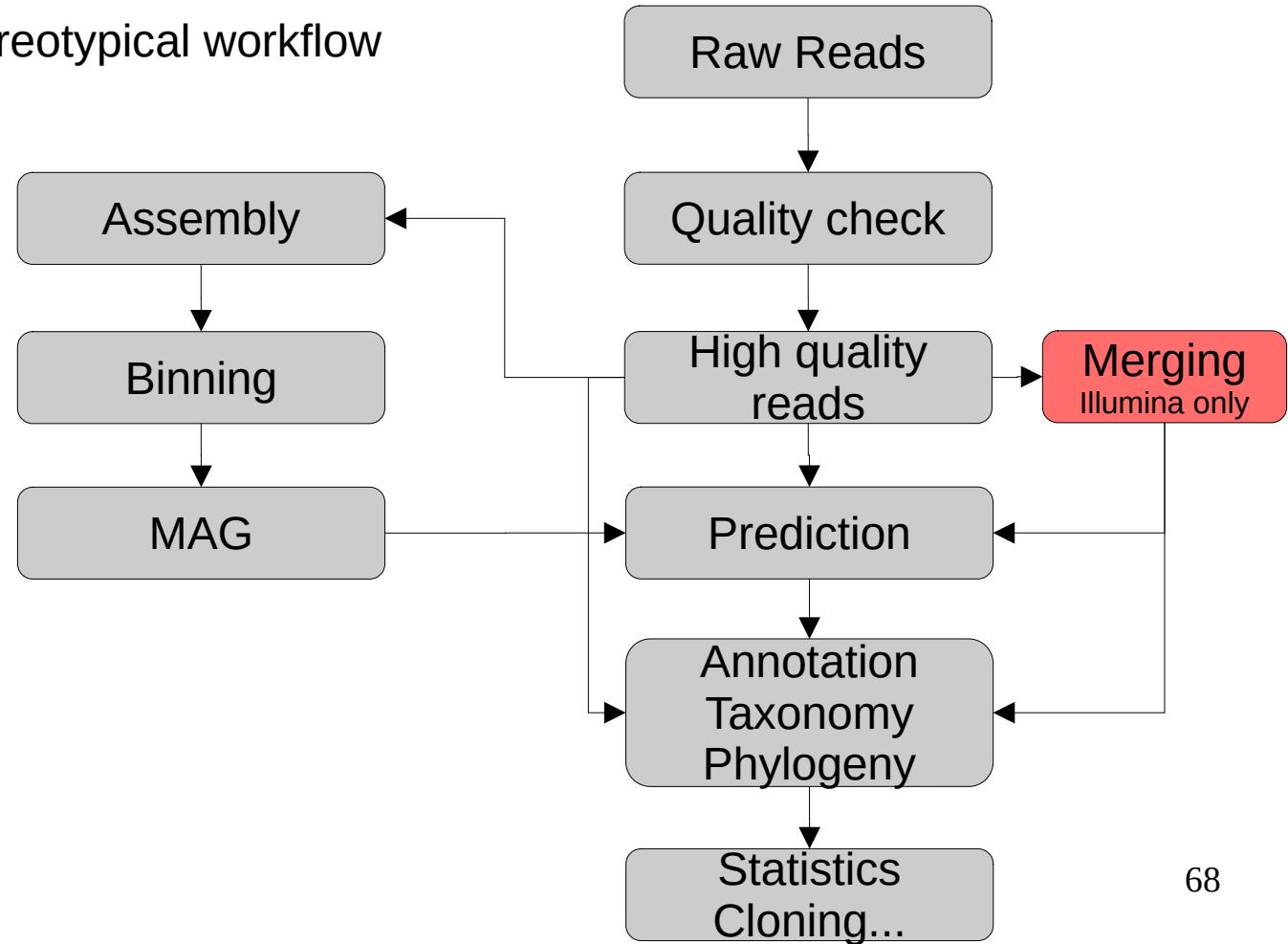
Phred+33 Quality

Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Bioinformatic analysis

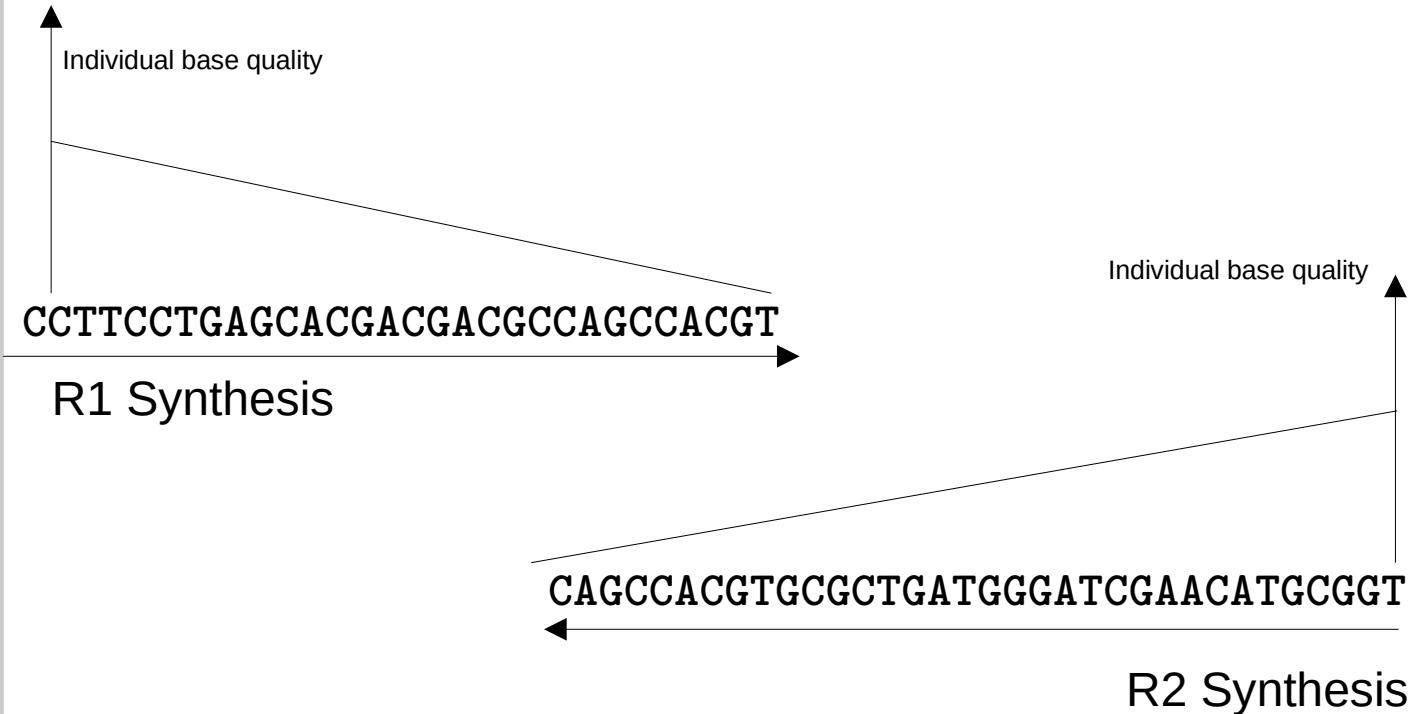
Stereotypical workflow



Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

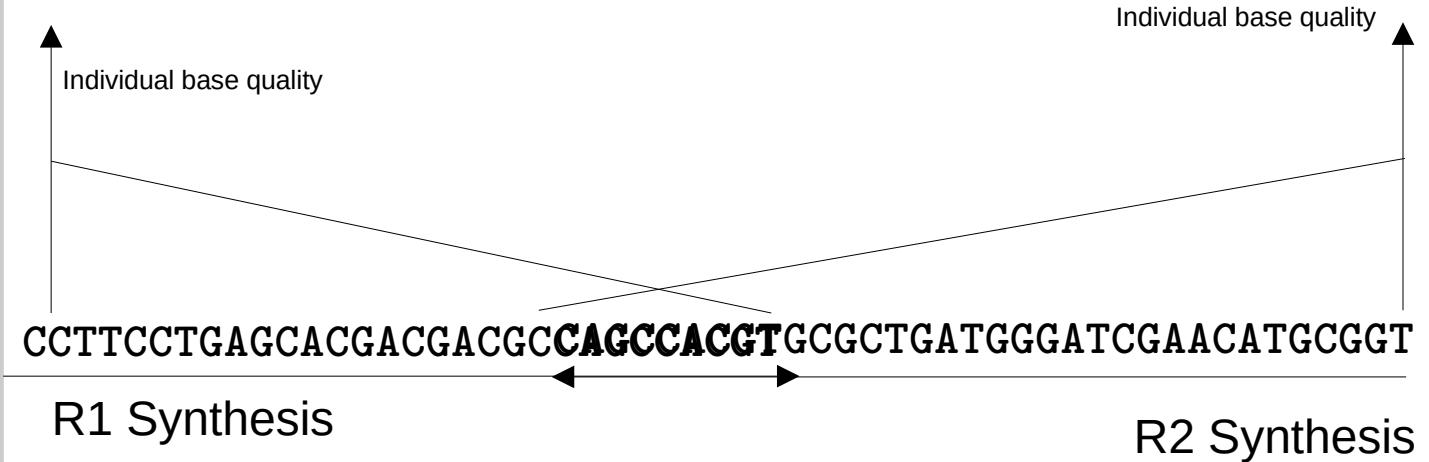
I – Metagenomics : Illumina Pair-end reads R1-R2 merging



Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Illumina Pair-end reads R1-R2 merging

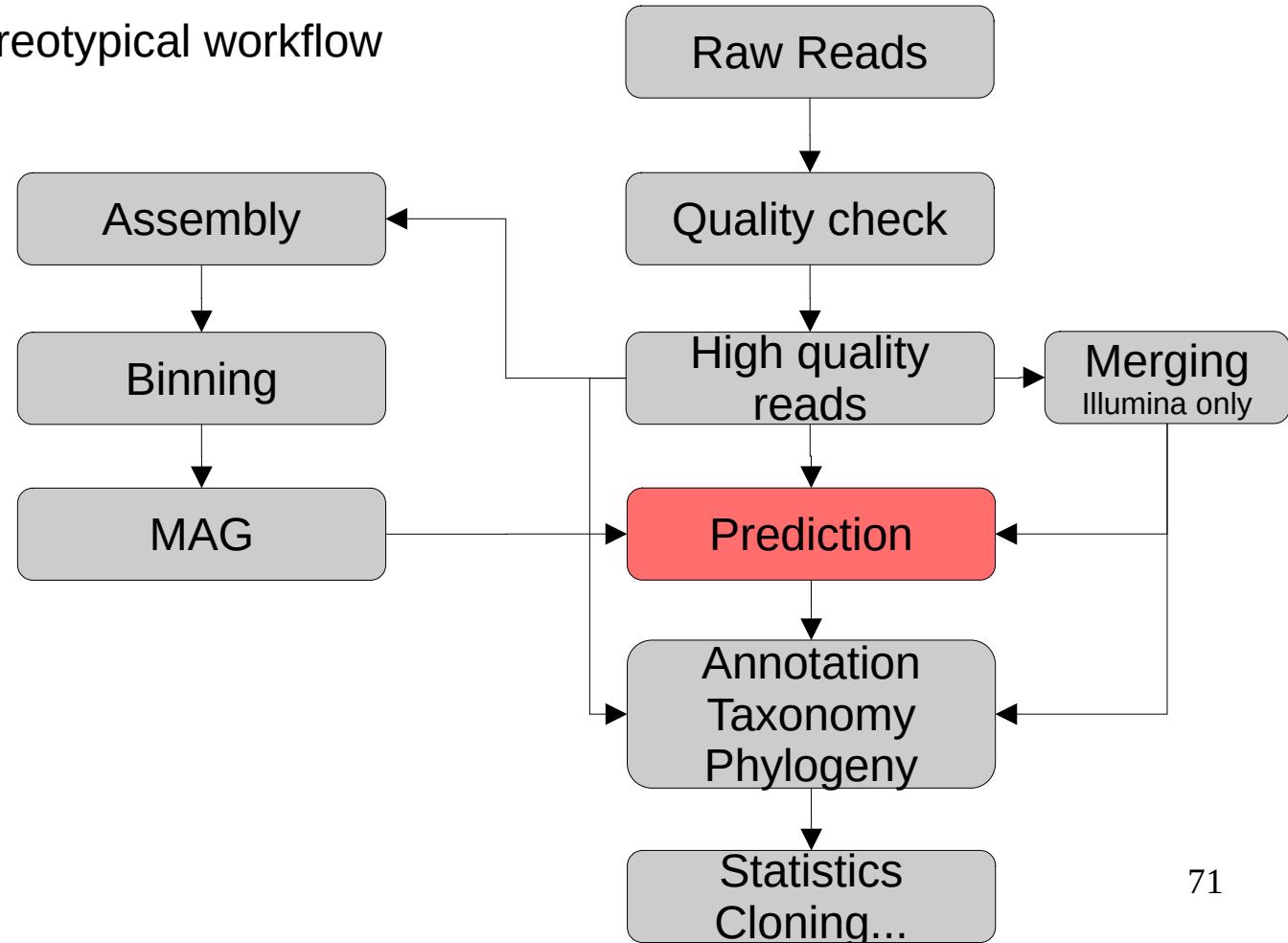


Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Bioinformatic analysis

Stereotypical workflow





Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Prediction

TTCGGTAAGCTCGATCTGCCCGCGCACTCCAGCGACTGGGCTTCCCTGCCACTACGAA
AATGCGATCTCATAACAGAATTGAGTTCAATGTCGTCTCTCTGGTAGTAGCCAGCCT
GCCAAGAACCGAGATCCACGTTCGCTCGCGCAATTCGCGATATTGATCTGCCAACCGCGC
TGCATGCCTCATATGGACGGTGATACCCGGATACC GGCGACGAAGTTGCTGAAGCACGGT
TGAGAGCAGGCCGGCGATAATGGCCTCATTGCCTCCAATGCTGATTCCCGAGCGTCGGG
TCGCTCCAGAAACTCTATATTTCACCCCTGACGCAAATCGTCAAACACCGCCGCCGC
GCAGTGGAGCAGCGCGACCATGGGTGGTCGGTCAACACCTGGCGGCTACGATCCAA
GAGACGCACGCCAATTGCGCGTTCTAGTTCTGCGATTGCCCGAGAAACCGCGGCTGGGT
CGTGGAGGCCGCTGCCCGCGCCTTCCATGCTGCCGGATTGAGCCACGGTCGTGAGA
ACGTAGAGGTCTGCAGCTTGAGTCGTCCGACAAATTCCGAAAATCCATAGCCAAC
CCATCTCAACATTGATATGACGACATGCGATAATAGCATTATGGCTTATGGCATCATCA
AACATCCGTCGGTGAACGGCTTAAACGATCGCAATGTGAACACTACAGGGAGAACG
CCGTCCCAAACCGAGGCATCCTGTTGTAACTGGTTGTCGACCGCTCGGTGGAAAGGGCT
TCTGCCCAATCTCAGCGACCGCTCAGTGGCTGATTAGCGATCCAGCGGGGCAGATGGC
CTAACTGGAGCCGTGCATCCGACCACCCACAAACACACATGGGAACCGTAATGAATATG
CGCAAGGACGATCAATTGTCGGCGCGCCGCGGAAACGAACGACGACGCGTGCCTCTGGA
TCGCTCTCAGCGCTCGAAAAAGCTGGGATCGAAGCCGCCCTCAATAGGTGTCGTGCCCTC

Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Prediction

TTCGGTAAGCTCGATCTGCCCGCGCACTCCAGCGACTGGGCTTCCCTGCCACTACGAA
A**ATG**CGATCTTCATACAGAATTGAGTTCAATGTCGTCTCTGGTAGTAGCCAGCCT
GCCAAGAACGAGATCCACGTTCGCTCGCGCAATTCGCGATATTGATCTGCCAACCGCGC
TGCATGCCATATGGACGGTGATAACCGGATACC GGCGACGAAGTTGCTGAAGCACGGT
TGAGAGCAGGCCGGCGATAATGGCCTCATGCCCTCAATGCTGATTCCCAGCGTCGGG
TCGCTCCAGAAACTCTATATTTCACCCCTGACGCAAATCGTCAAACACCGCCGCCGC
GCAGTGGAGCAGCGCGACCATGGGTGGTCGGTCAACCTGGCGGCTACGATCCAA
GAGACGCACGCAATTGCGCGTTCTAGTTCTGCATTGCCGAGAAACCGCGGGCTGGGT
CGTGGAGGCCGCTGCCGCCCTTCCATGCTGCCGGATTGAGCCACGGTCGTGAGA
ACGTAGAGGTCTGCAGCTTGAGTCGTCCGACAAATTCCGAAAATCCATAGCCAAC
CCATCTCAACATTGATATGACGACATGCATAATAGCATTATGGCTTATGGCATCATCA
AACATCCGTCGGTGAACGGCTTAA

		Second letter					
		U	C	A	G		
First letter	U	UUU UUC UUA UUG } Phe	UCU UCC UCA UCG } Ser	UAU UAC UAA Stop UAG Stop } Tyr	UGU UGC } Cys	U C A G }	
	C	CUU CUC CUA CUG } Leu	CCU CCC CCA CCG } Pro	CAU CAC CAA CAG } His	CGU CGC CGA CGG } Arg	U C A G }	
	A	AUU AUC AUA AUG } Ile	ACU ACC ACA ACG } Thr	AAU AAC AAA AAG } Asn	AGU AGC AGA AGG } Ser	U C A G }	
	G	GUU GUC GUA GUG } Val	GCU GCC GCA GCG } Ala	GAU GAC GAA GAG } Asp	GGU GGC GGA GGG } Gly	U C A G }	

- Introduction
- **Metagenomics**
 - Origins
 - Methodologies**
 - Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

What can be predicted ?

- Introduction
- Metagenomics
 - Origins
 - Methodologies
 - Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

What can be predicted ?

CDS

tRNA

rRNA

Signal Peptide

Intermembrane domain

Specific sites (Cleaving, anchoring ...)

CRISPR palindromes

Restriction enzyme sites ...

- Introduction
- Metagenomics
 - Origins
 - Methodologies
 - Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Prediction and Annotation

How are can we be sure ?

CDS and other motifs are blasted against databases

It's more efficient than blasting like a idiot

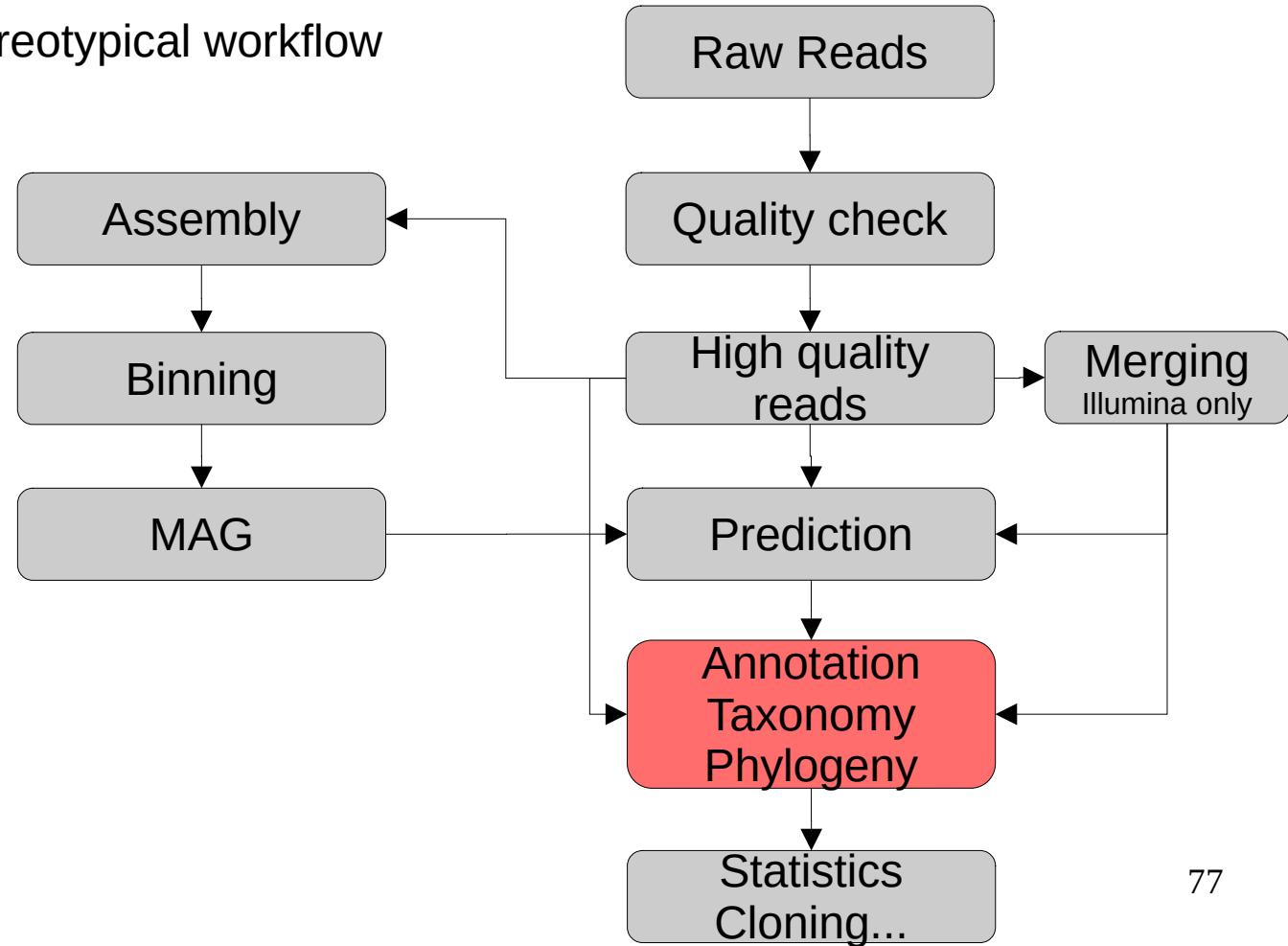


Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Bioinformatic analysis

Stereotypical workflow



- Introduction
- Metagenomics
 - Origins
 - Methodologies
 - Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

16836	16363	CDS	<pre>inference ab initio prediction:Prodigal:002006 locus_tag NEIDAFCF_00019 product hypothetical protein</pre>
17408	16995	CDS	<pre>EC_number 2.7.13.3 db_xref COG:COG0642 gene phoQ_1 inference ab initio prediction:Prodigal:002006 inference similar to AA sequence:UniProtKB:P23837 locus_tag NEIDAFCF_00020 product Sensor protein PhoQ</pre>
18422	17592	CDS	<pre>inference ab initio prediction:Prodigal:002006 locus_tag NEIDAFCF_00021 product hypothetical protein</pre>
19122	18397	CDS	<pre>gene phoP_1 inference ab initio prediction:Prodigal:002006 inference similar to AA sequence:UniProtKB:P0DM78 locus_tag NEIDAFCF_00022 product Virulence transcriptional regulatory protein PhoP</pre>
19385	19122	CDS	<pre>inference ab initio prediction:Prodigal:002006 locus_tag NEIDAFCF_00023 product hypothetical protein</pre>
19477	20010	CDS	

I – Metagenomics : Specific annotations

- Introduction
- **Metagenomics**
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

Specific annotations can be carried out with specific databases
Like Antibiotic resistance database etc

- Introduction
- **Metagenomics**
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

Who's there ?

- Introduction
- **Metagenomics**

 - Origins
 - Methodologies

 - Productions

 - Applications
 - Database
 - Limits and Perspectives
 - Conclusion

Target 16S rRNA DNA coding
Sequences

Metagenomics

- Introduction
- Metagenomics
 - Origins
 - Methodologies
 - Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Taxonomy – Phylogeny

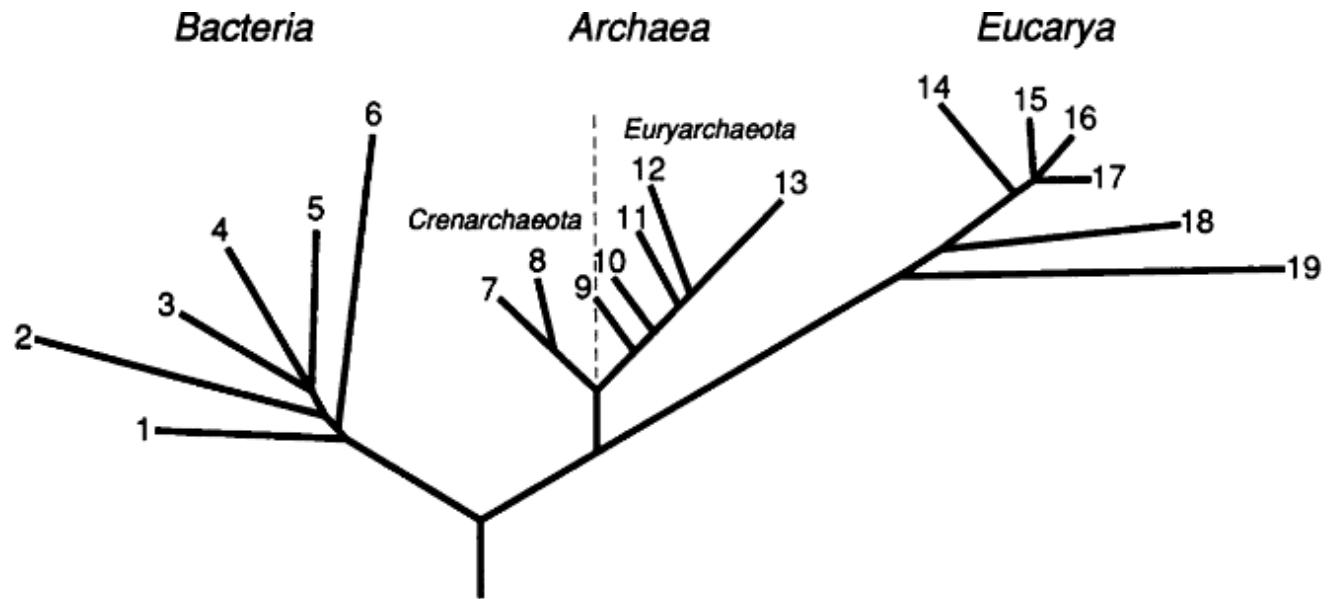
Target 16S rRNA DNA coding Sequences



Metabarcoding like

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

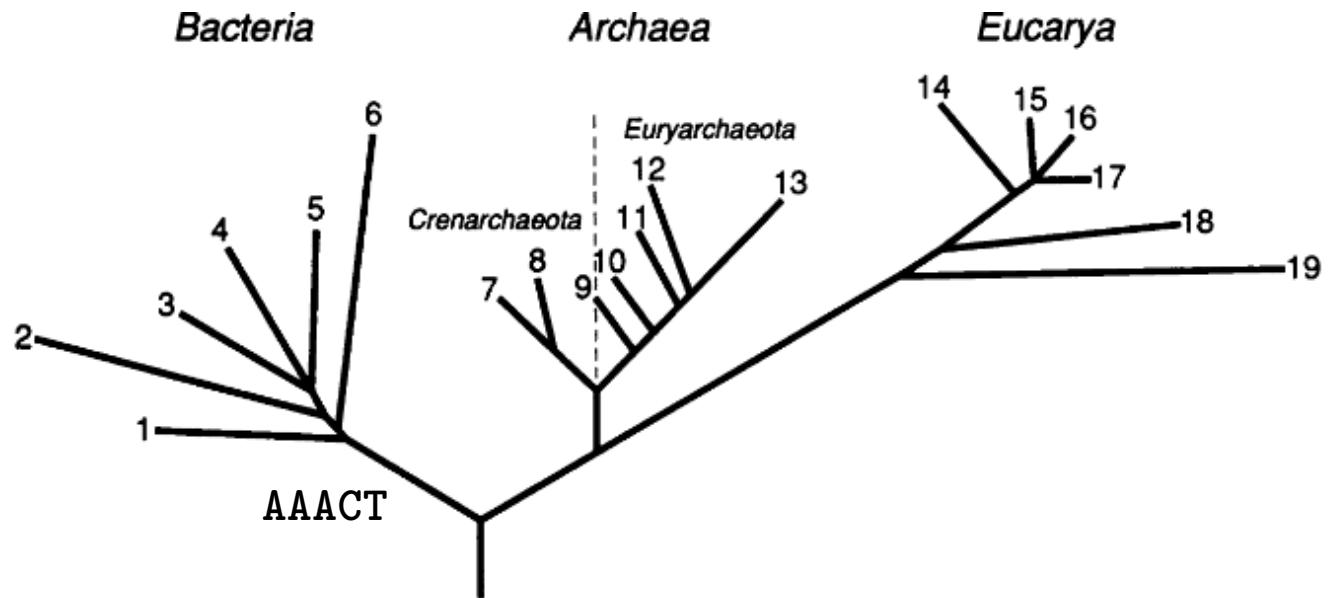
Kmer counting



Ex : Kraken2, Concoct, Centrifuge ...

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

Kmer counting



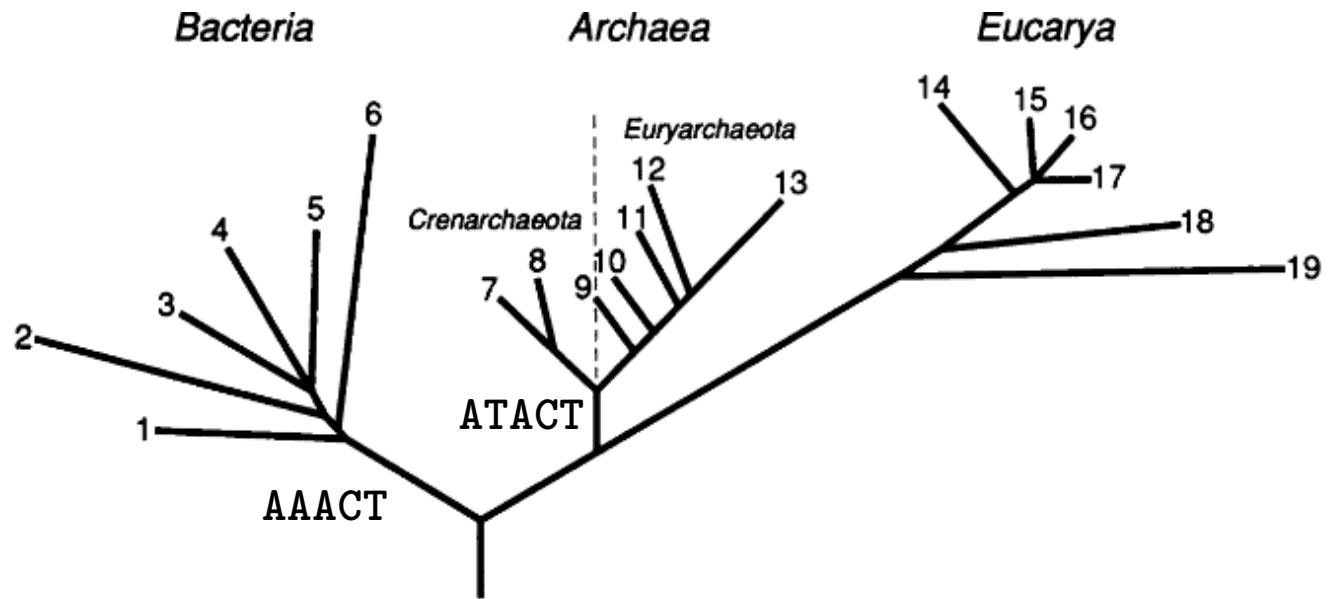
Ex : Kraken2, Concoct, Centrifuge ...

Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Taxonomy – Phylogeny

Kmer counting



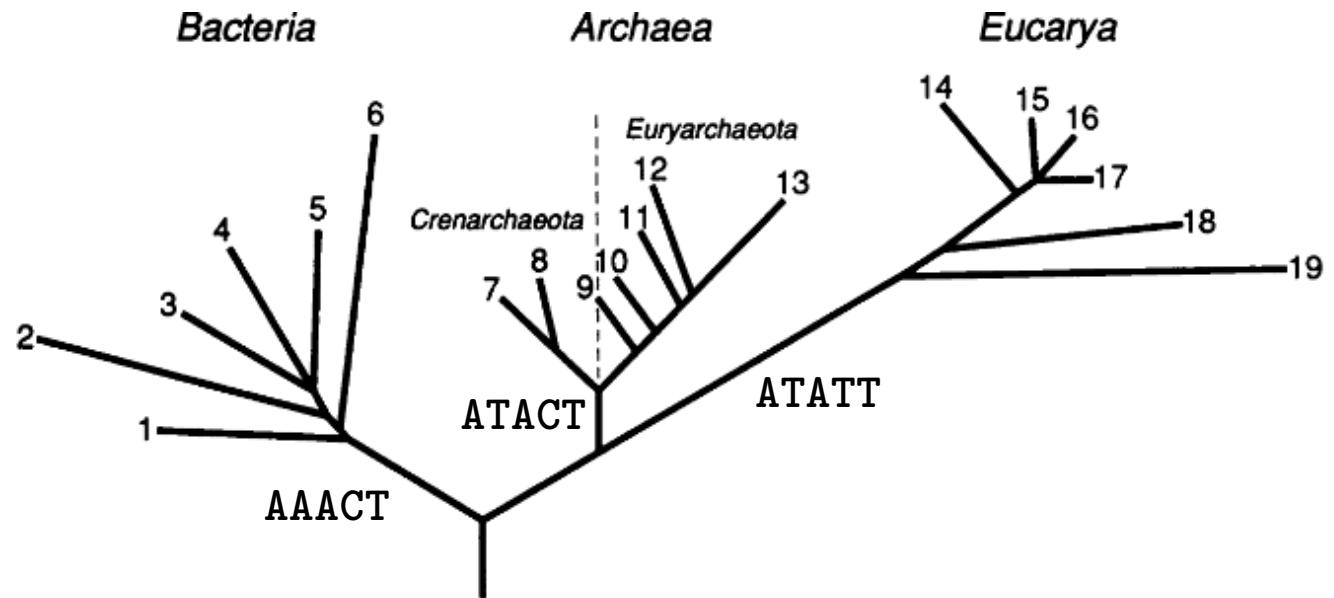
Ex : Kraken2, Concoct, Centrifuge ...

Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Taxonomy – Phylogeny

Kmer counting



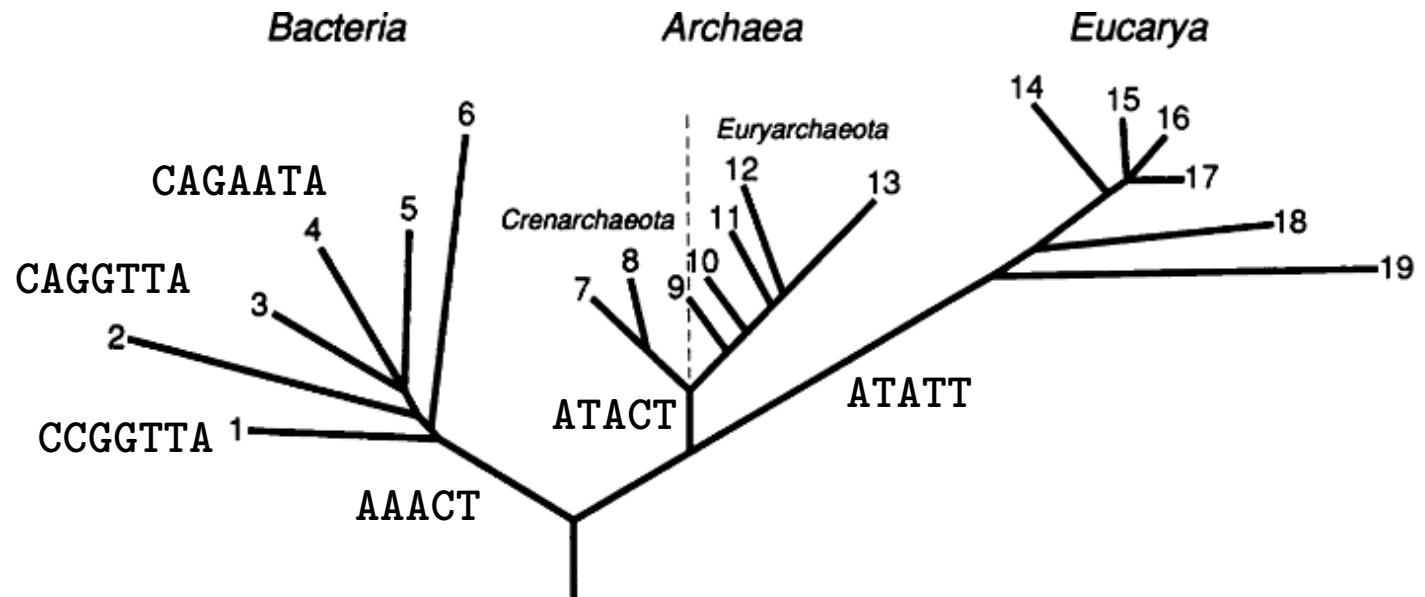
Ex : Kraken2, Concoct, Centrifuge ...

Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Taxonomy – Phylogeny

Kmer counting



Ex : Kraken2, Concoct, Centrifuge ...

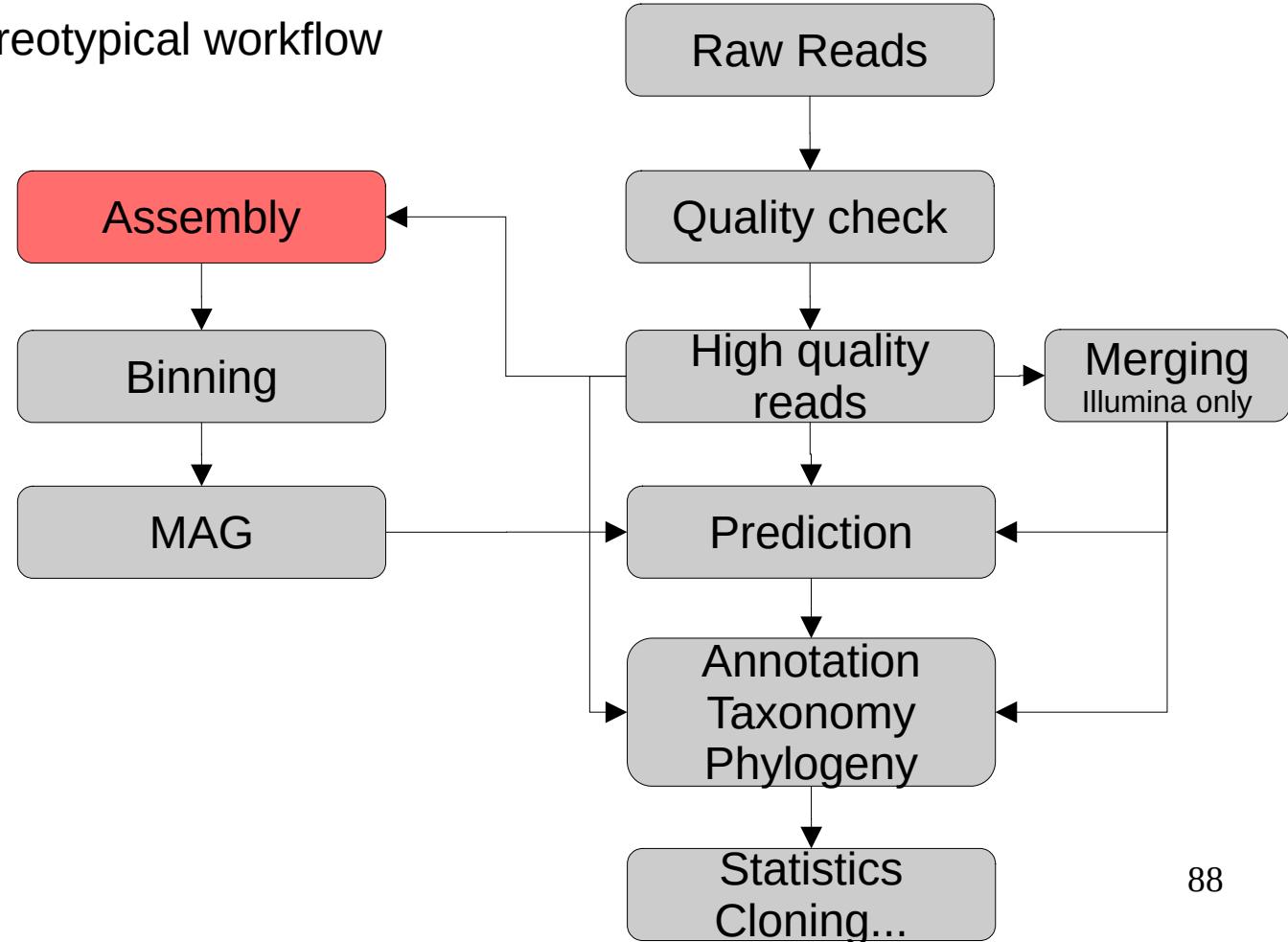


Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Bioinformatic analysis

Stereotypical workflow



- Introduction
- **Metagenomics**
 - Origins
 - Methodologies**
 - Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion



Metagenomics

- Introduction
- Metagenomics
 - Origins
 - Methodologies
 - Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Assembly

CCTTCCTGAGCACGACGACGCCAGCCACGT

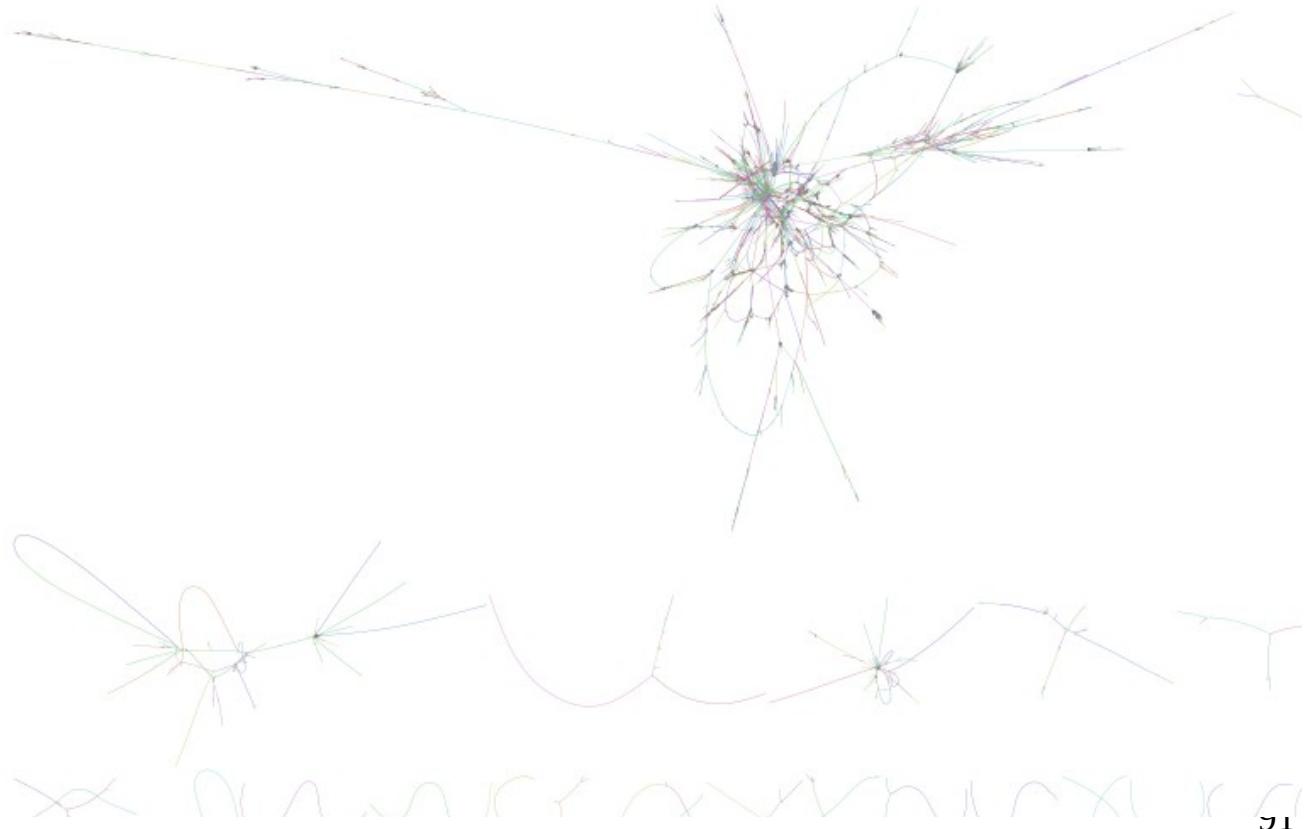
CAGCCACGTGCGCTGATGGGATCGAACATGCGGT

Flye, Medaka, Unicycler, Megahit etc.

Metagenomics

- Introduction
- **Metagenomics**
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

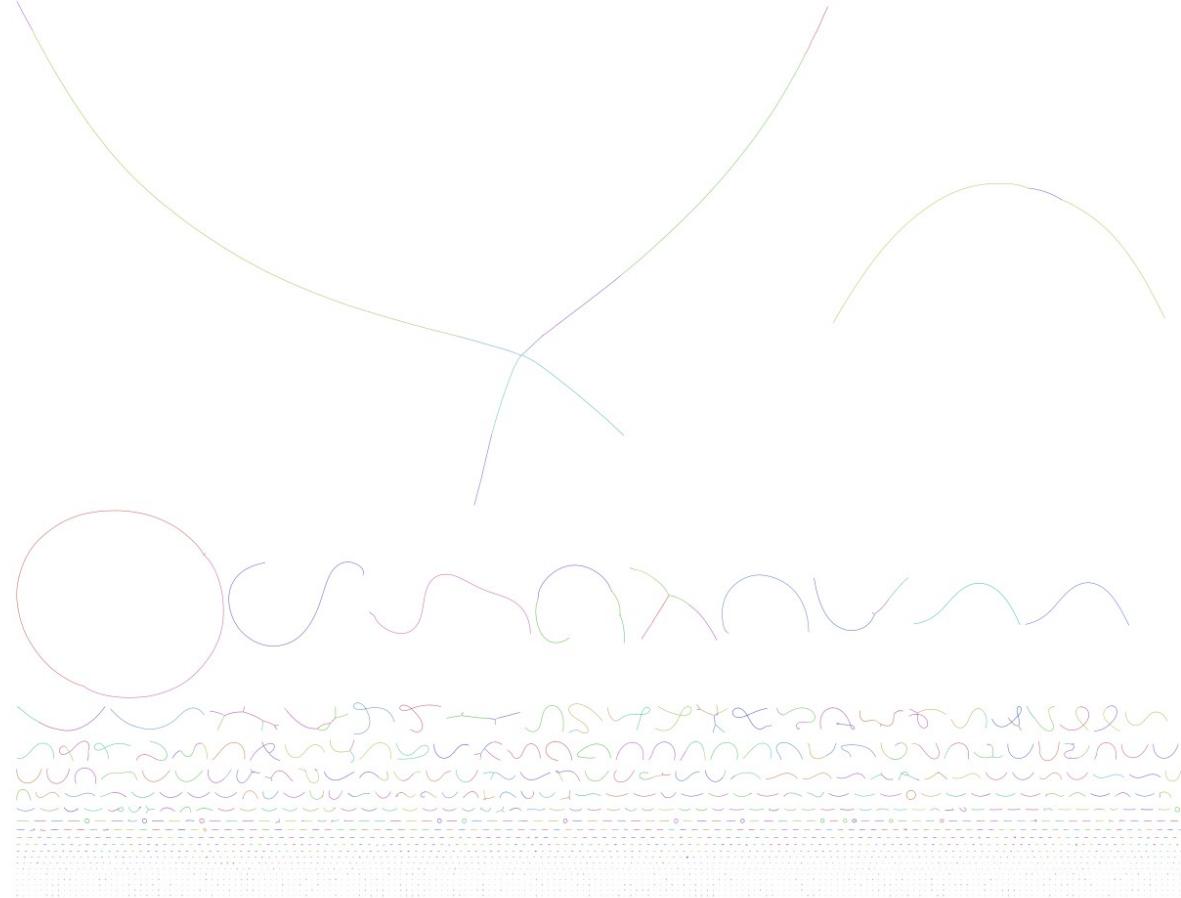
I – Metagenomics : Assembly



Metagenomics

- Introduction
- **Metagenomics**
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Assembly

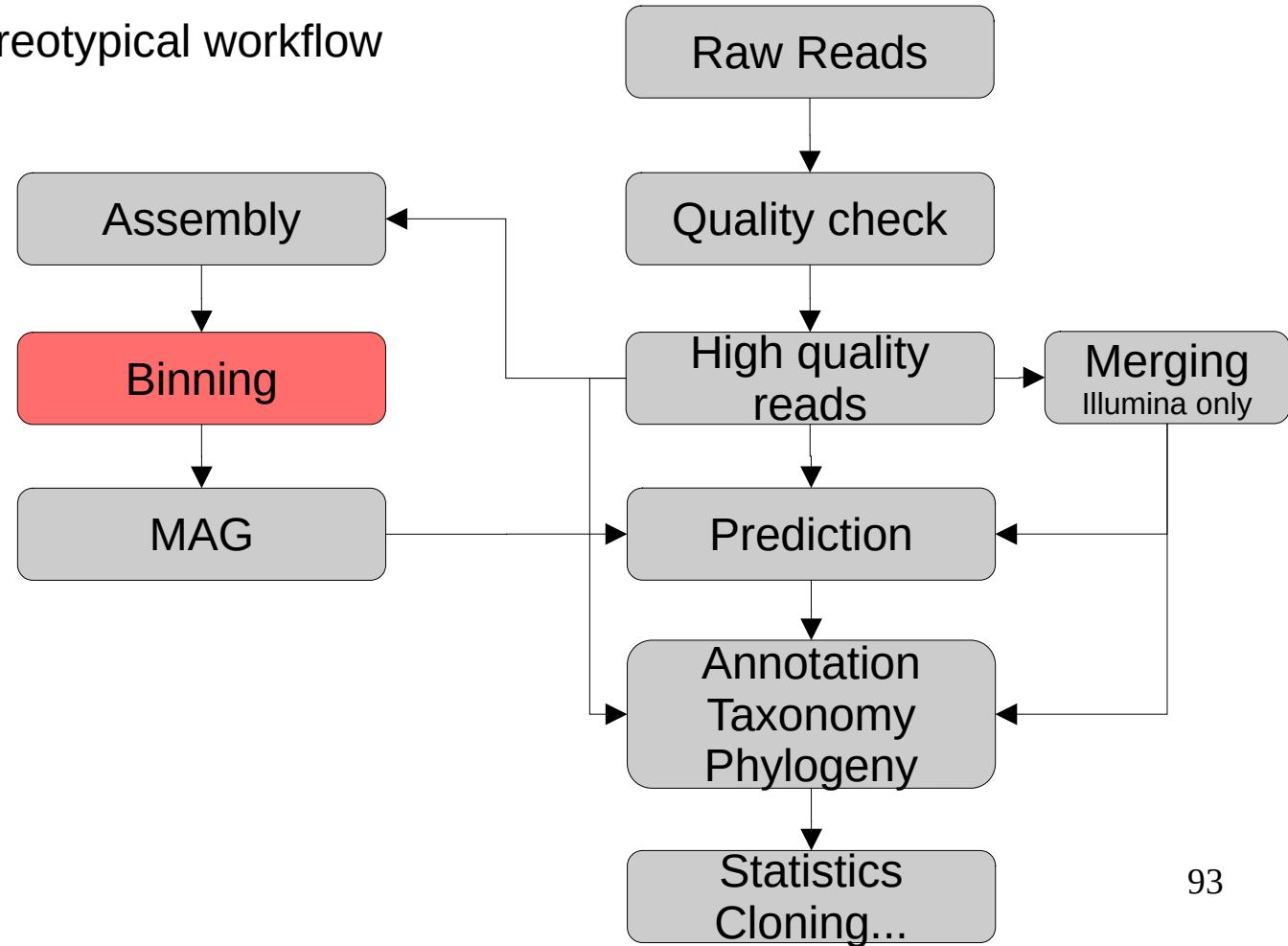


Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Bioinformatic analysis

Stereotypical workflow



I – Metagenomics : Binning

- Introduction
- **Metagenomics**
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

When pieces don't stick together.

How can we know that they belong to the same organism ?

- Introduction
- **Metagenomics**
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

Kmer frequencies

Codon usage bias

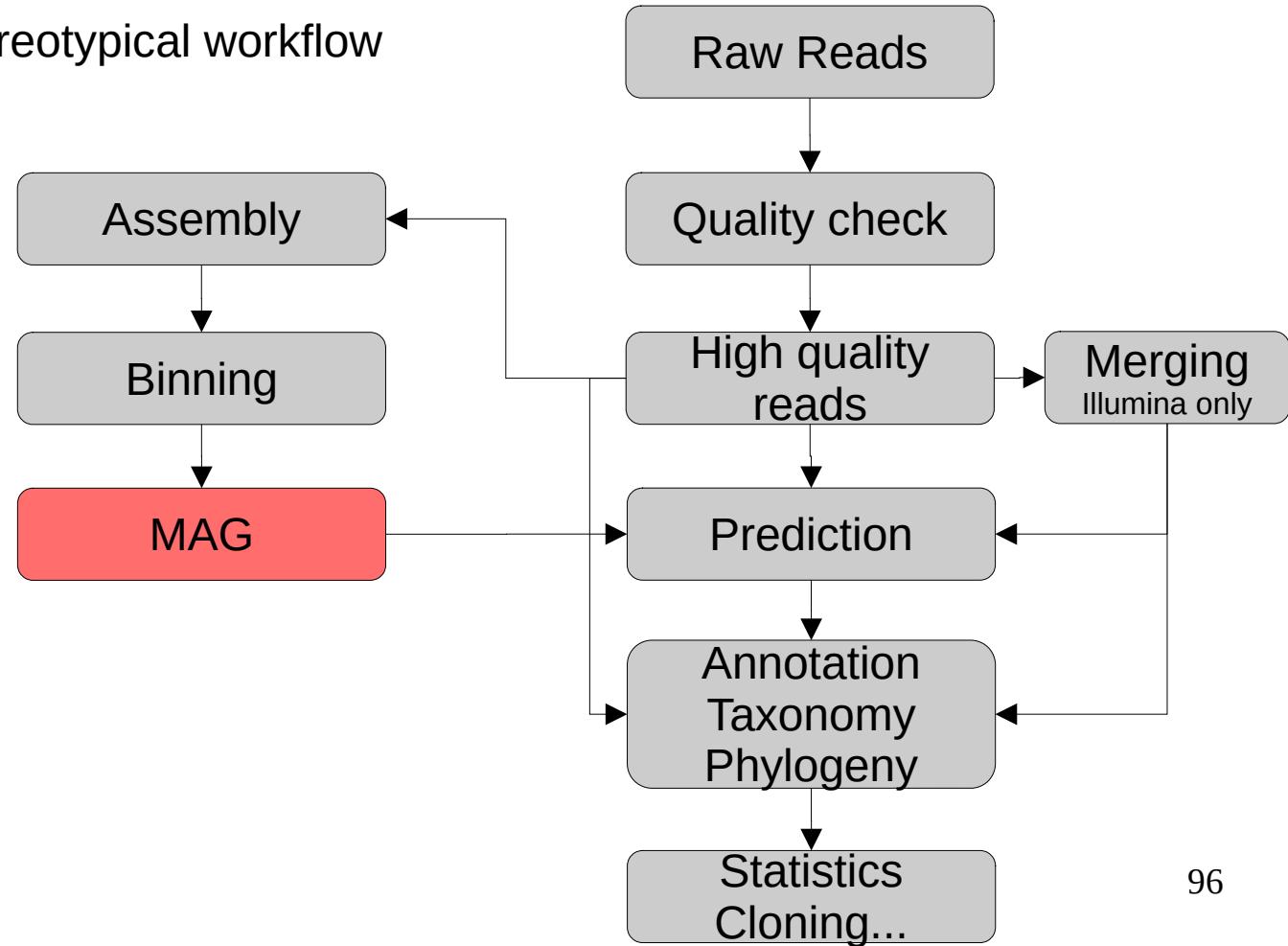
GC content

Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Bioinformatic analysis

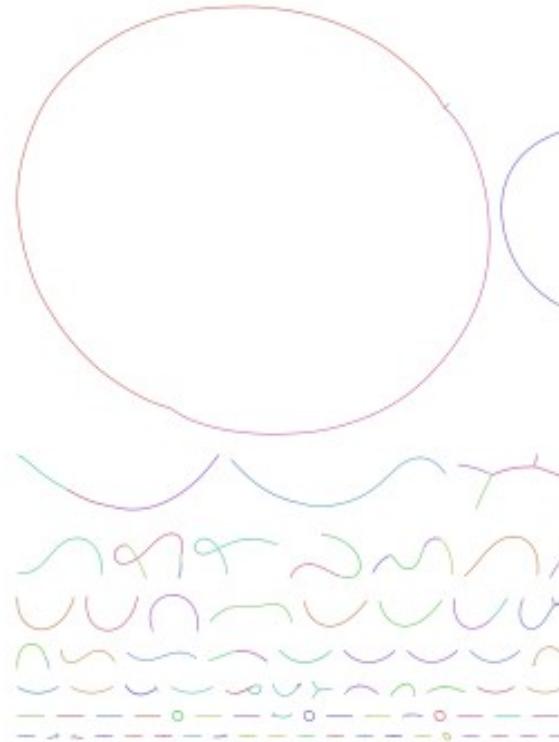
Stereotypical workflow



Metagenomics

- Introduction
- **Metagenomics**
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : MAGs



A Metagenome-Assembled Genome (MAG) is a genome reconstructed from metagenomic data.

- Introduction
- Metagenomics
 - Origins
 - Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

I – Metagenomics : Other productions

Industrial enzymes
Antibiotic discoveries

...

A wide-angle photograph of a rural landscape. In the foreground, there's a field of tall, yellowish-green grass. Beyond it, a series of green hills roll across the scene. A small, dark body of water, possibly a pond or a small lake, is nestled among the trees on one of the hills. The sky above is a clear blue with scattered white clouds.

Finding our way in and out
of the forest

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

You can do metagenomics on anything.

We focus on bacteria, but it works for anything

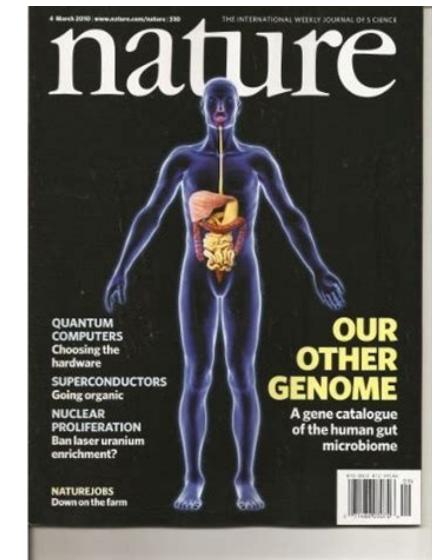
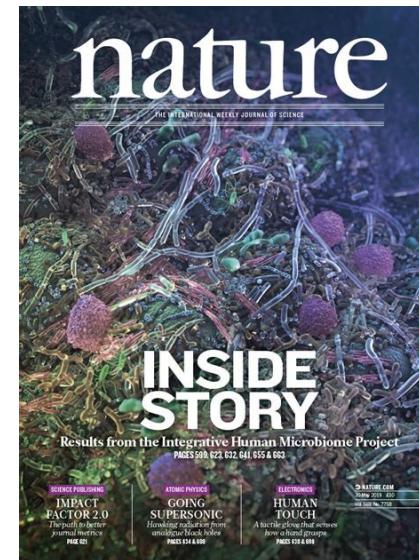
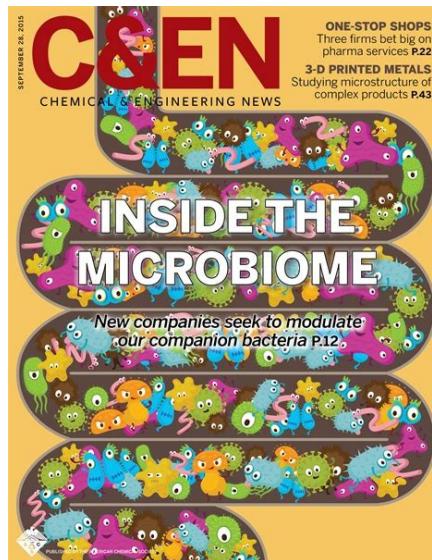
Ancient DNA is metagenomics

Sewage, Soil, deep waters, sediments, spatial samples, meteorites, hydrothermal vents ...

Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

II – Applications : Examples



- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- **Applications**
- Database
- Limits and Perspectives
- Conclusion



- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

II – Applications : Examples



ELSEVIER

Available online at www.sciencedirect.com



Atmospheric Environment 39 (2005) 4143–4153

**ATMOSPHERIC
ENVIRONMENT**

www.elsevier.com/locate/atmosenv

Microbial population in cloud water at the Puy de Dôme: Implications for the chemistry of clouds

Pierre Amato^{a,b,c}, Matthieu Ménager^a, Martine Sancelme^a, Paolo Laj^b,
Gilles Mailhot^c, Anne-Marie Delort^{a,*}

^aLaboratoire de Synthèse et Etude de Systèmes à Intérêt Biologique, UMR 6504 CNRS-Université Blaise Pascal,
63177 Aubière Cedex, France

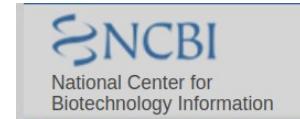
^bLaboratoire de Météorologie Physique, UMR 6016 CNRS-Université Blaise Pascal, 63177 Aubière Cedex, France

^cLaboratoire de Photochimie Moléculaire et Macromoléculaire, UMR 6505 CNRS- Université Blaise Pascal,
63177 Aubière Cedex, France

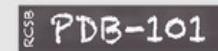
Received 3 December 2004; accepted 8 April 2005

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

III – Database



GREENGENES



- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

In regione caecorum rex est luscus

Au royaume des aveugles les borgnes sont rois

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- **Limits and Perspectives**
- Conclusion

Reconstruction is computer intensive

Hard to be sure of anything

You see what you know

Biases in databases will have impact on analysis

Are you sure that your threshold makes you see what you want to see ?

Did you cross-contaminate ?

...

Conclusion

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

Metagenomics is the science arising from everything known before
(Maybe like all sciences)

It allows exciting raw discoveries

This is a new way to see the world around us

In metagenomes might lie the secret of Bacteria

Metagenomics

- Introduction
- Metagenomics
- Origins
- Methodologies
- Productions
- Applications
- Database
- Limits and Perspectives
- Conclusion

Thank you for your attention

More ressources :

<https://merenlab.org/2020/07/27/history-of-metagenomics/>

<https://merenlab.org/2020/07/01/dark-side/>

