

Phylogenetic Annotation of Next generation AMplicons

ePANAM

<http://panam-meb.univ-bpclermont.fr/panam.php>

Phylogenetic Affiliation of SSU rRNA Genes Generated by Massively Parallel Sequencing: New Insights into the Freshwater Protist Diversity

Najwa Taib^{1,2}, Jean-François Mangot^{1,2,3,4}, Isabelle Domaizon^{3,4}, Gisèle Bronner^{1,2}, Didier Debroas^{1,2*}

¹ Clermont Université, Université Blaise-Pascal, Laboratoire "Microorganismes: Génome et Environnement", BP 10448, Clermont-Ferrand, France, ² CNRS, UMR 6023, LMGE, Aubiere, France, ³ INRA, UMR 42 CARRTEL, Thonon les Bains, France, ⁴ Université de Savoie, UMR 42 CARRTEL, Le Bourget du Lac, France

Abstract

Recent advances in next-generation sequencing (NGS) technologies spur progress in determining the microbial diversity in various ecosystems by highlighting, for example, the rare biosphere. Currently, high-throughput pyrotag sequencing of PCR-amplified SSU rRNA gene regions is mainly used to characterize bacterial and archaeal communities, and rarely to characterize protist communities. In addition, although taxonomic assessment through phylotag is considered as the most

- ◆ Analyse complète de la diversité
- ◆ Jeux de données volumineux(Pyrotags, MiSeq)
- ◆ 16S et 18S ARNr
- ◆ Implémente une approche phylogénétique ; décrit les clades
- ◆ Calcule les indices de diversité phylogénétique

ePANAM
Depicting the microbial diversity from high-throughput amplicons

Home • New Analysis • About • FAQ • Contact • Register • Login

version 1.0 | last modified : Sept, 0th 2013 | admin : panam_AT_univ-bpclermont.fr

454 Roche **Illumina MiSeq** Choose your datatype

*** All the fields should be filled.
Do not forget your email address so you will be notified when your results are available.

Project metadata
Name:
Comments:

Data files

F/R1 sequences Aucun fichier sélectionné.
R/R2 sequences Aucun fichier sélectionné.
Tags Aucun fichier sélectionné.

[download barcode example file](#) each file size may not exceed 400MB

Domains

Eukaryota
Bacteria
Archaea

Traitement des données issues des NGS : PANAM

```
Use -gamma for approximate but comparable Gamma(20) log-likelihoods
ML-NNI round 2: LogLk = -81755.250 NNIs 92 max delta 42.55 Time 38.92es (max delta 42.546)
ML-NNI round 3: LogLk = -81176.762 NNIs 63 max delta 70.93 Time 44.12es (max delta 59.209)
ML-NNI round 4: LogLk = -80965.872 NNIs 41 max delta 27.10 Time 47.94es (max delta 27.101)
ML-NNI round 5: LogLk = -80820.670 NNIs 24 max delta 33.02 Time 51.06es (max delta 33.017)
ML-NNI round 6: LogLk = -80737.619 NNIs 12 max delta 18.13 Time 53.32es (max delta 18.127)
ML-NNI round 7: LogLk = -80679.707 NNIs 13 max delta 22.36 Time 54.96
ML-NNI round 8: LogLk = -80494.091 NNIs 10 max delta 51.58 Time 56.34
ML-NNI round 9: LogLk = -80422.407 NNIs 5 max delta 16.75 Time 57.26
ML-NNI round 10: LogLk = -80419.810 NNIs 0 max delta 0.00 Time 58.08
Turning off heuristics for final round of ML NNIs (converged)
ML-NNI round 11: LogLk = -80222.356 NNIs 4 max delta 1.26 Time 65.26 (final) delta 1.256)
Optimize all lengths: LogLk = -80216.857 Time 67.03
Total time: 73.65 seconds Unique: 341/345 Bad splits: 1/338 Worst delta-LogLk 0.459
panam@bioinfo1:~$ █
```

Programmation

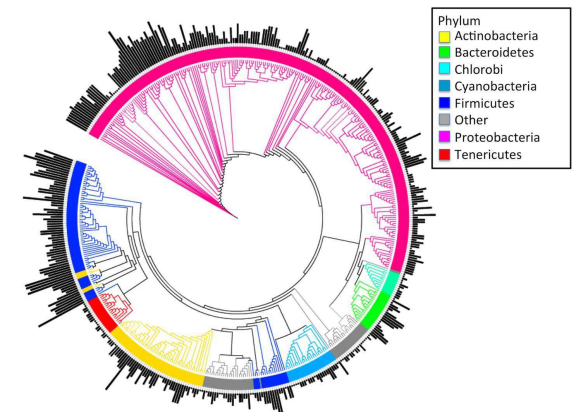
Séquences brutes



```
>seq1
TAGTGGTCATCCAACGTGTGAAC
>seq2
GCTATCGCCCGATGAGCCATGTAA
>seq3
GGTCNNNCCATGCATTTTTTTTTT
>seq4
ATTAGTGGTCATCCACTAGAGATC
>seq5
GCTATNNNNNNNNNNATGAGNNN
```

PANAM

Structure des communautés



Traitement des données issues des NGS : PANAM

Phylogenetic Annotation of Next Generation Amplicons

Prétraitement des séquences

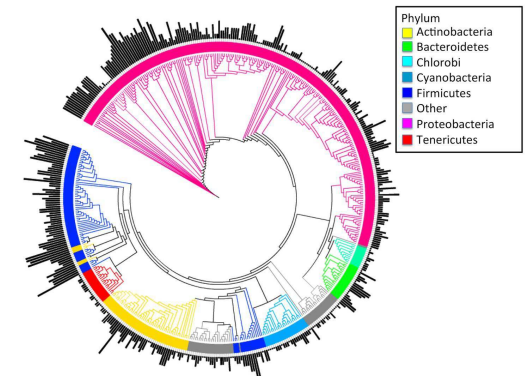
Séquences brutes



```
>seq1
TAGTGGTCATCCAACACTGTGTAAC
>seq2
GCTATCGCCCGATGAGCCATGTAA
>seq3
GGTCNNNCCATGCATTTTTTTTTT
>seq4
ATTAGTGGTCATCCACTAGAGATC
>seq5
GCTATNNNNNNNNNNNATGAGNNN
```

```
>seq1
TAGTGGTCATCCAACACTGTGTAAC
>seq2
GCTATCGCCCGATGAGCCATGTAA
>seq3
GGTCNNNCCATGCATTTTTTTTTT
>seq4
ATTAGTGGTCATCCACTAGAGATC
>seq5
GCTATNNNNNNNNNNNATGAGNNN
```

Structure des communautés



Traitement des données issues des NGS : PANAM

Phylogenetic Annotation of Next Generation Amplicons

Séquences brutes



```
>seq1
TAGTGGTCATCCAACACTGTGTAAC
>seq2
GCTATCGCCCGATGAGCCATGTAA
>seq3
GGTCNNNCCATGCATTTTTTTTTT
>seq4
ATTAGTGGTCATCCACTAGAGATC
>seq5
GCTATNNNNNNNNNNNATGAGNNN
```

Prétraitement des séquences

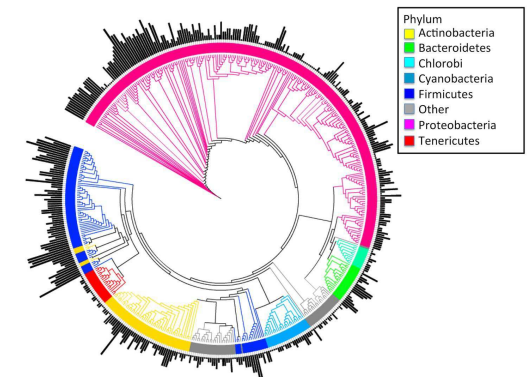
Identification des individus

```
seq1
Bacteria, Proteobacteria, Gamma Proteobacteria,
Enterobacteriales, Enterobacteriaceae,
Escherichia, Escherichia coli
```

```
seq2
Bacteria, Proteobacteria, Epsilon Proteobacteria,
Campylobacteriales, Helicobacteriaceae,
Helicobacter, Helicobacter pylori
```

```
seq4
Bacteria, Proteobacteria, Gamma Proteobacteria,
Enterobacteriales, Enterobacteriaceae,
Escherichia, Escherichia vulneris
```

Structure des communautés



Traitement des données issues des NGS : PANAM

Phylogenetic Annotation of Next Generation Amplicons

Séquences brutes



```

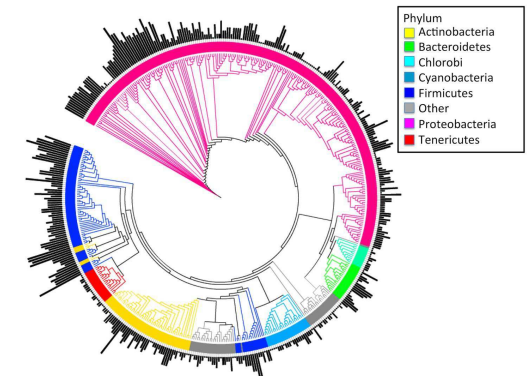
>seq1
TAGTGGTCATCCAACGTGTGTAAC
>seq2
GCTATCGCCCGATGAGCCATGTAA
>seq3
GGTCNNNCCATGCATTTTTTTTTT
>seq4
ATTAGTGGTCATCCACTAGAGATC
>seq5
GCTATNNNNNNNNNNNATGAGNNN
    
```

Prétraitement des séquences

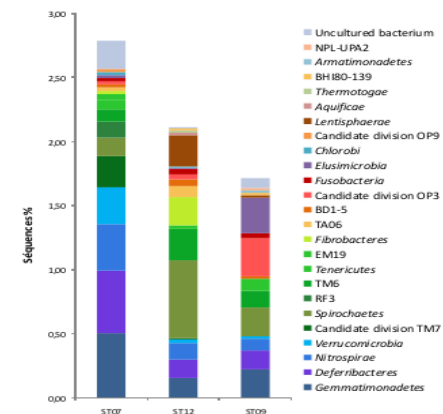
Identification des individus

Visualisation des résultats

Structure des communautés



ST12 (9555)	ST09 (5361)	total (19042)	Affiliation
19,59	13,52	14,83	Bacteria;Chloroflexi
3,94	0,4	2,98	Bacteria;Chloroflexi
3,18	5,15	3,69	Bacteria;Chloroflexi
0,2	1,64	1,08	Bacteria;Chloroflexi;Caldilineae;Caldilineales;Caldilineaceae;Caldilinea
0,3	1,42	1,02	Unclassified Bacteria
0,8	1,79	1,32	Unclassified Bacteria
0,0	2,37	1,11	Bacteria;Firmicutes;Clostridia;Clostridiales;Syntrophomonadaceae
0,4	2,24	1,22	Unclassified Bacteria
0,3	1,75	1,01	Bacteria;Proteobacteria
0,4	3,00	1,27	Bacteria;Proteobacteria;Deltaproteobacteria
2,34	0,1	1,32	Bacteria;Cyanobacteria
31,5	33,4	30,8	



Identification des espèces : Annotation taxonomique

Séquence inconnue

GCTATCGCCCGATGAGC

Comparaison à l'existant



Identification des espèces : Annotation taxonomique

Séquence inconnue

GCTATCGCCCGATGAGC

Comparaison à l'existant



Probabilités

Similitude

Evolution

mothur

QIIME
Quantitative Insights Into Microbial Ecology

MG-RAST
metagenomics analysis server

PANAM

Identification des espèces : Annotation taxonomique

Séquence inconnue

GCTATCGCCCGATGAGC

Comparaison à l'existant



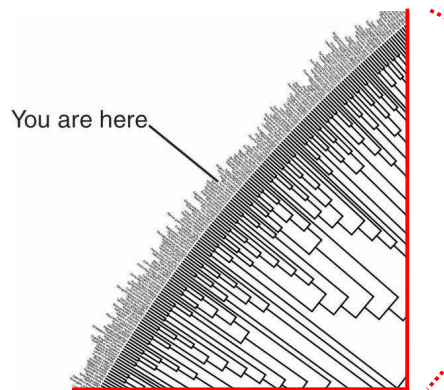
Probabilités

Similitude

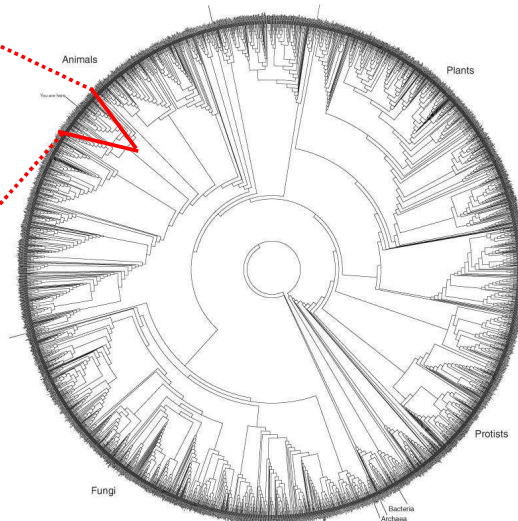
Evolution

Histoire évolutive

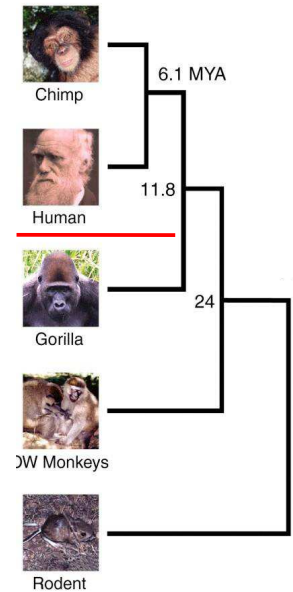
Liens de parenté



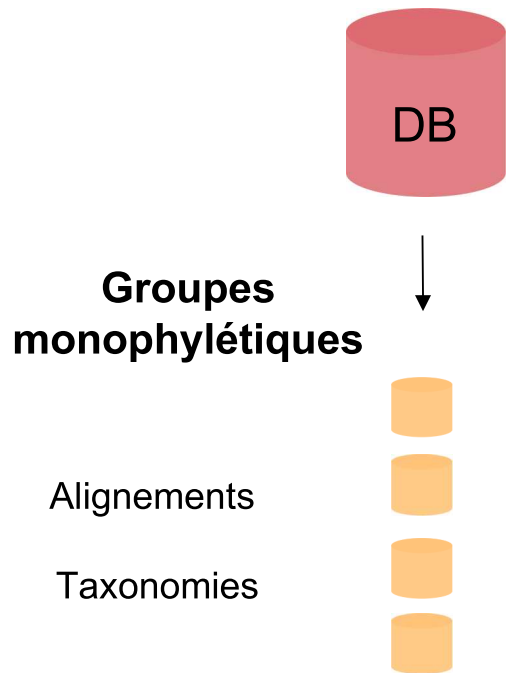
Groupe phylogénétique



Phylogénie du monde vivant

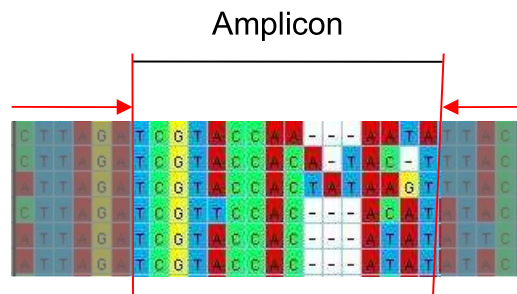


PANAM : Etapes d'affiliation

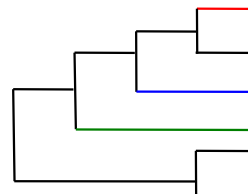


Comparaison

Alignements de profils



Reconstruction phylogénétique



NGS reads

Reads

```
>F16CHBQ04JDAWQ
GCCTCCC TCGCGCATCAGTGCCGCGC
GGTAATTCAGCTCCAATAGCGTATATTAA
AGTTGTTGCAAGTAAAAAGCTCGTAGTTG
AATTTGGCCTGGTTGGGTGGTATGCGC
CGCAAGGTCATTACTGCTTGACCCGGGT
>F1CQM UO02JPNSW
TGCCCTCCC TCGCGCCATCAGTGCCGCGC
CGGTAATTCAGCTCCAAGTGCGGTATATT
AAAGTTGTTGCGGTTAAAAGCTCGTAGTT
GGATTTCTGCCGAGGACGACCGATCGAC
CCATACGGGTTTGTATCTGGT TAGACCGA
```

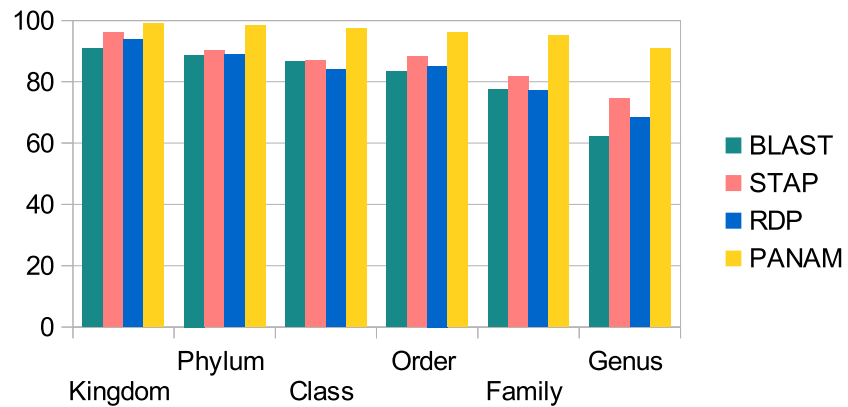


Tri des reads

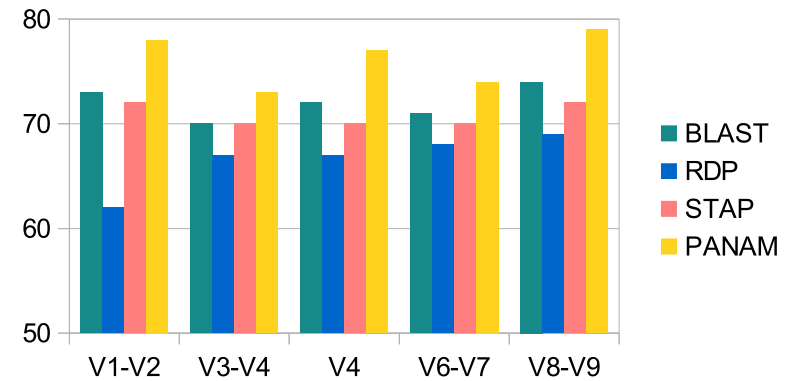
Restitution taxonomique

Séquences de référence

Séquences complètes



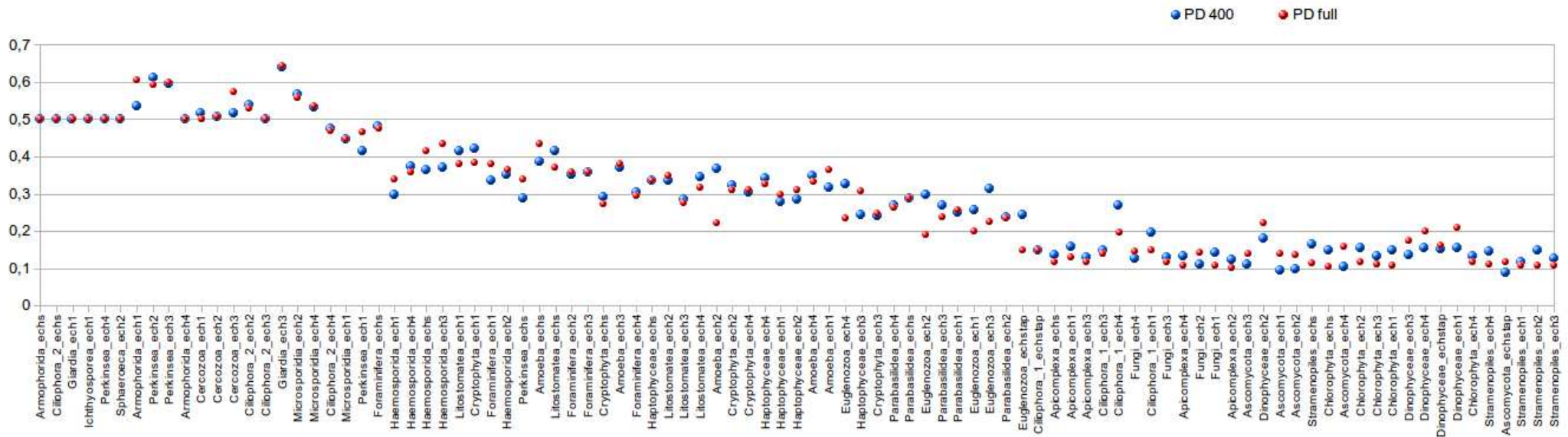
Amplicons de 400 pb



Indices de diversité phylogénétique

Séquences de référence

Indice PD sur des phylogénies de séquences complètes et des phylogénies de 400 pb



Test Friedman : $p > 5\%$

Pas de différence significative des indices PD calculés à partir de phylogénies de séquences complètes et d'amplicons de 400 pb

Restitution phylogénétique

Séquences environnementales

Topologies des arbres

Pas de différence significative des topologies des arbres générés à partir de séquences complètes et de 400 pb (Mantel test $p < 5\%$)

La position sur le SSU rRNA

Le groupe monophylétique

Description des clades

