

Rapport de laboratoire

Initiation aux NGS

Robin COHEN, Xavier TOMISINEC, Guillaume TRICKNOT

05/05/2024

MA1 Section Ingénieur en technologie du
vivant - HEH

Département Science et Technologique

Année académique 2023-2024

Table des matières

1	Introduction.....	1
2	Principe	1
2.1	PCR.....	1
2.2	Illumina – Miseq.....	1
2.3	Analyse métagénomique.....	1
3	Manipulation réalisée.....	2
3.1	Échantillons	2
3.2	Extraction et purification de l'ADN	2
3.3	Normalisation de la concentration en ADN	3
3.4	Amplicon PCR	3
3.5	Purification Post-PCR	4
3.6	Index PCR	4
3.7	Quantification et normalisation de concentration des librairies	4
3.8	Dénaturation et lancement du séquençage	5
4	Résultats et interprétation.....	6
4.1	Analyse réalisée de novo	6
4.1.1	Analyse qualité.....	6
4.1.2	Trimming.....	7
4.1.3	Taxonomie.....	8
4.2	Analyse avec les données réalisées avec les données du pipeline	9
4.2.1	Analyse des phylums.....	9
4.2.2	AAanalyse des genres	12
4.2.3	Calcul de l'indice de diversité.....	21
4.3	Comparaison des résultats	23
4.3.1	Comparaison Compost.....	23
4.3.2	Comparaison Eau	25
4.3.3	Comparaison Sol	28
4.3.4	Comparaison Jus de patate	30
4.4	Interprétation des résultats.....	32
4.4.1	Compost.....	32
4.4.2	Eau	32
4.4.3	Terre.....	32
4.4.4	Jus de patate.....	33
5	Conclusion	34

6	Bibliographie	35
7	Annexe.....	36
7.1	FastQC	36
7.1.1	FASTQC Compost.....	36
7.1.2	FASTQC Guillaume Eau.....	42
7.1.3	FASTQC Guillaume Terre	48
7.1.4	FASTQC Jus de patate.....	54
7.2	Krona	60
7.2.1	Compost.....	60
7.2.2	Eau	61
7.2.3	Terre.....	63
7.2.4	Jus de patate.....	64
7.3	Pavian.....	65
7.3.1	Composte.....	65
7.3.2	Eau	66
7.3.3	Terre.....	67
7.3.4	Jus de patate.....	69

1 Introduction

Lors de la semaine dédiée à l'initiation du séquençage nouvelle génération, nous avons abordé les principes théoriques et pratiques du séquençage, avec un accent particulier sur la technologie Illumina ainsi que sur le workflow de l'Illumina MiSeq. Ce rapport abordera les principes théoriques, le schéma analytique global de la procédure ainsi qu'une analyse et une interprétation des différents résultats obtenus.

2 Principe

2.1 PCR

La PCR (Polymerase Chain Reaction) est une technique permettant d'amplifier de manière exponentielle un matériel génétique en fournissant les primer requises, les 4 dNTP, et une ADN polymérase. Elle fonctionne sous forme de cycle composé d'étapes de :

- Dénaturation : Séparation des deux brins d'ADN à haute température
- Hybridation : Fixation de courtes séquences d'ADN complémentaires (amorces) aux extrémités du fragment d'ADN à amplifier
- Élongation : Synthèse d'un nouveau brin d'ADN par l'enzyme ADN polymérase à partir des amorces

2.2 Illumina – Miseq

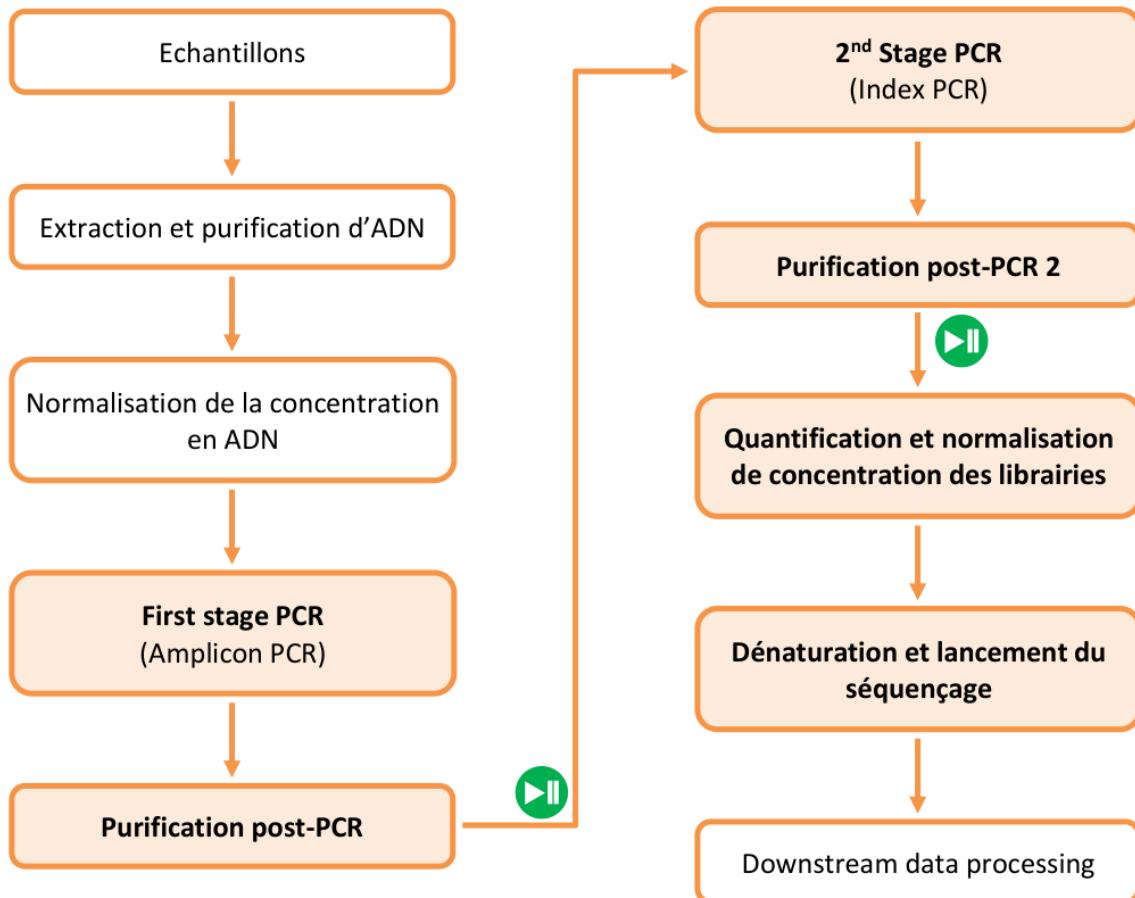
MiSeq est un séquenceur de nouvelle génération de la société Illumina. La machine a pour particularité d'utiliser la technologie de séquençage par synthèse chimique (SBS) pour déterminer la séquence d'acides nucléiques d'un échantillon d'ADN. Le séquenceur a besoin de fragment d'ADN et utilise une bridge PCR pour créer des clusters denses de fragments d'ADN sur une surface solide (flow cell).

2.3 Analyse métagénomique

La métagénomique est une approche de l'étude génomique qui permet d'étudier l'ensemble du matériel génomique présent dans un environnement pour offrir une vue sur l'ensemble de la communauté microbienne d'un environnement.

3 Manipulation réalisée

Pour commencer, voici les différentes étapes que nous avons suivies lors de notre laboratoire.



▶ : Cold Storage Option

Figure 1 : Schéma général suivi durant le laboratoire

3.1 Échantillons

Nous avons eu la possibilité de ramener des échantillons d'origine environnementale. Les différents échantillons de notre groupe sont les suivants : un échantillon de terre de jardinage, un prélèvement d'eau de la rivière (l'eau d'heure), d'un compost de jardin et d'une pomme de terre pourrie.

3.2 Extraction et purification de l'ADN

L'extraction d'ADN s'est fait à l'aide de kits d'extraction dédiés au différent type d'échantillon. Pour préparer nos échantillons à l'extraction, les échantillons terreux ont été passés au tamis afin d'éliminer d'éventuels éléments non biologiques et les échantillons liquides ont été centrifugé et les surnageant éliminé. Ensuite, nous avons utilisé un kit d'extraction d'ADN DNeasy PowerSoil (1) permettant d'extraire et purifier de l'ADN. Ce kit entraîne une lyse par réaction chimique et par homogénéisation mécanique permettant ainsi d'enlever les inhibiteurs présents dans le sol pouvant altérer un processus ultérieur. L'échantillon lysé est ensuite purifié par lavage en étant fixé sur une membrane de silice.



Figure 2 Extraction d'ADN d'un échantillon de terre

3.3 Normalisation de la concentration en ADN

À la suite de l'extraction de l'ADN, nous avons vérifié la pureté et la concentration de l'ADN pour chaque échantillon.

Au niveau de la normalisation, cette étape consiste à standardiser la quantité d'ADN de chaque échantillon à un niveau optimal afin d'assurer une amplification uniforme et efficace lors de la PCR.

La concentration initiale d'ADN de chaque échantillon est mesurée en utilisant la technique du NanoDrop. Cette méthode nous a fourni une quantité d'ADN en $\mu\text{g/mL}$ ce qui permet ensuite d'évaluer la dilution à réaliser afin d'être dans la gamme de mesure qui était de $30\text{ng}/\mu\text{L}$ maximum.

Voici un tableau reprenant les différentes dilutions ainsi que les différentes valeurs de pureté de l'ADN :

Échantillon	[] ($\mu\text{g/mL}$)	Volume final (μL)	Volume Échantillon (μL)	Volume H ₂ O (μL)	Pureté (260/280) de l'échantillon
Eau Guillaume	1	15	Pas de dilution	/	1,97
Sol Guillaume	50	15	9	6	2
Jus de Patate Xavier	12	15	Pas de dilution	/	1,95
Composte Robin	163,5	15	2,8	12,2	1,97

Tableau 1 : Tableau des dilutions et pureté de l'ADN

Basés sur la pureté, nous pouvons dire que tous les échantillons possèdent une contamination ARN.

3.4 Amplicon PCR

Lors de cette étape, nous avons réalisé l'amplification par PCR à l'aide du Phusion Plus Master Mix. Ce protocole inclut la préparation du master mix et la distribution de ce dernier dans les puits de la plaque de réaction. Le cycle PCR commence par une dénaturation initiale à 98°C pendant 30 secondes, suivie de cycles de dénaturation, d'hybridation (température ajustée en fonction des amorces) et d'extension. Une extraction finale est ensuite réalisée.

Une fois l'amplification terminée, les produits PCR sont vérifiés par électrophorèse sur gel d'agarose pour s'assurer de leur taille et leur qualité. Une fois validés, ils sont ensuite purifiés en utilisant soit les kits *AMPure XP Beads* ou *SparQ PureMag Beads*.

3.5 Purification Post-PCR

La purification post-PCR se fait grâce à l'utilisation de billes magnétiques qui exploitent leur capacité à se lier sélectivement aux acides nucléiques présents dans le mélange réactionnel. Une fois ajoutées à l'échantillon contenant les produits de PCR, ces billes magnétiques capturent spécifiquement les fragments d'ADN amplifiés ainsi que les contaminants tels que les amorces, les dNTPs et les polymérases non désirées. En appliquant un champ magnétique, les billes avec les acides nucléiques liés sont rassemblées au fond du tube ou du récipient, ce qui permet de facilement éliminer les contaminants non liés dans le surnageant. Une fois le surnageant retiré, les billes avec les acides nucléiques purifiés peuvent être récupérées en suspendant les billes dans un tampon adapté. Ce processus offre une purification hautement efficace, améliorant la qualité des échantillons et la fiabilité des résultats ultérieurs.

3.6 Index PCR

Cette étape consiste à ajouter les index et les adaptateurs nécessaires pour le séquençage par PCR aux amplicons. Les index ont été ajoutés à l'aide du *Phusion Plus Master Mix* dans une PCR de 8 cycles.

L'insertion des index ont ensuite été vérifiée par électrophorèse sur gel d'agarose.

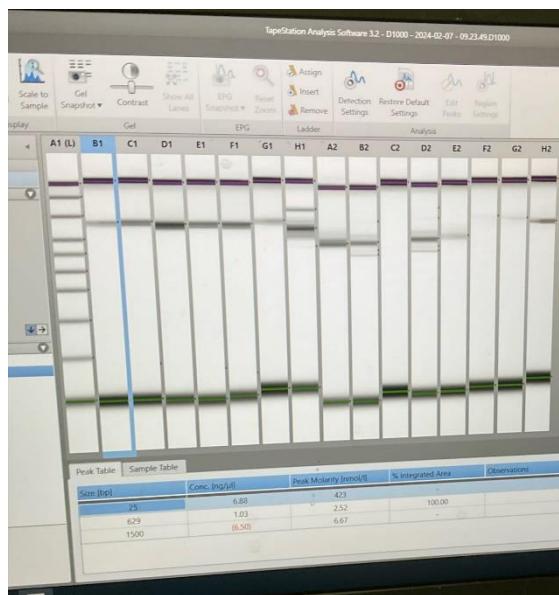


Figure 3 : Vérification des index

Les résultats ont montré une mauvaise insertion des librairies en C2, F2 et G2.

Les produits sont ensuite purifiés à l'aide de bille magnétique.

3.7 Quantification et normalisation de concentration des librairies

Pour déterminer la concentration et la pureté des échantillons d'ADN préalablement amplifiés, nous avons utilisé une machine *Nanodrop* combinée avec le *PicoGreen*. Le *Nanodrop* utilise la spectrophotométrie UV-visible pour mesurer l'absorbance de l'échantillon à différentes longueurs d'onde. Cette méthode permet d'estimer la concentration d'ADN en se basant sur l'absorbance à 260 nm, tandis que la pureté est évaluée en comparant le rapport entre l'absorbance à 260 nm et celle à

280 nm. Parallèlement, le *PicoGreen* est un colorant fluorescent qui se lie spécifiquement à l'ADN double brin, permettant une quantification précise de faibles concentrations d'ADN. Les concentrations obtenues nous ont permis de déterminer les dilutions à effectuer pour chaque extrait pour obtenir une concentration de 4 nM avec du RSB (*ReSuspension Buffer*) ou de 10mM avec du Tris.

Voici le tableau réalisé avec Excel afin d'avoir les volumes de librairie et de Tris10mM à prélever pour avoir un volume de normalisation de 20µL.

Échantillons	Volume de librairie à prélever (µL)	Volume de Tris10mM à ajouter (µL)	Molarité des librairies (nM)
Eau 16S	28,9	0	3
Sol 16S	15,7	4,3	5
Jus de patate 16S	10,3	9,7	8
Composte 16S	19,4	0,6	4
Eau ITS	10,7	9,3	7
Sol ITS	14	6	6
Jus de Patate ITS	17,6	2,4	5
Compost ITS	17,5	2,5	5
Eau 18S	/	0	0
Sol 18S	/	0	0
Jus de patate 18S	18	2	4
Composte 18S	67,1	0	1

Tableau 2 : Volume de librairies et de Tris à prélever

5µl de chaque librairie sont ensuite rassemblées dans 1 seul tube (pooling).

3.8 Dénaturation et lancement du séquençage

Une librairie basée sur le virus *PhiX* est ajoutée aux librairies *poolées*. Cela permet un apport d'une diversité de nucléotides afin d'assurer une Cluster Template Génération et une qualité de séquençage optimal.

Cette librairie est diluée avec du RSB ou du Tris afin de normaliser sa concentration aux autres librairies à 4nM.

Les librairies sont ensuite dénaturées au NaOH et par chauffage. Les librairies sont ensuite dénaturées une dernière fois à 4pM pour enfin être mélangées.

Une étape de nettoyage de la *flowcell* est ensuite nécessaire afin d'éliminer d'éventuels sels résiduels issus de son buffer de stockage.

Après une dernière étape de dénaturation par chauffage, les librairies peuvent être ajoutées sur la *flowcell* pour ensuite pouvoir charger la cartouche dans le *MiSeq* et lancer le séquençage.

4 Résultats et interprétation

4.1 Analyse réalisée de novo

4.1.1 Analyse qualité

Le programme *FastQC* est utilisé afin de vérifier la qualité des séquences obtenues. Deux graphiques sont intéressants à analyser, la qualité et le contenu des séquences. Dans l'idéal, la qualité doit être supérieure à 20 et le graphique du contenu des séquences doit être le plus linéaire possible. Les graphiques suivants représentent une tendance de toutes les autres données ressue.

Forward:

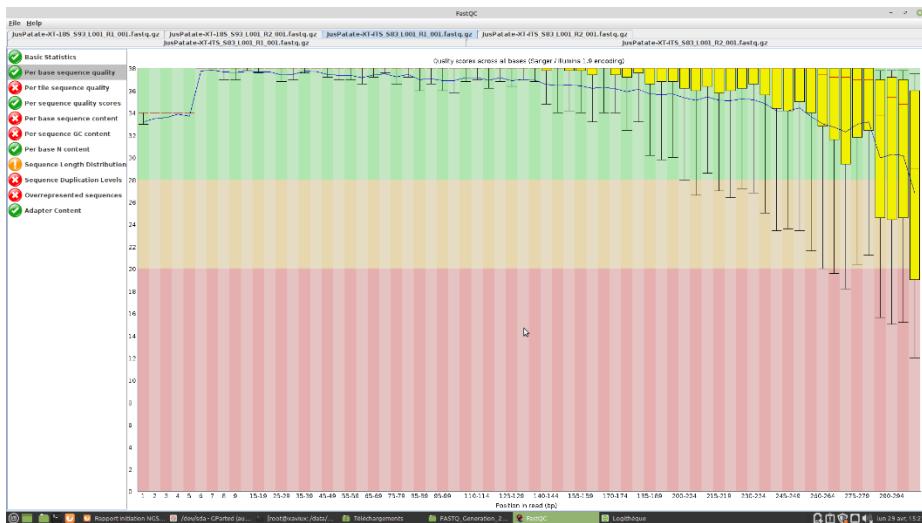


Figure 4 Résultat de l'analyse qualité par séquence de l'ITS du jus de patate en forward

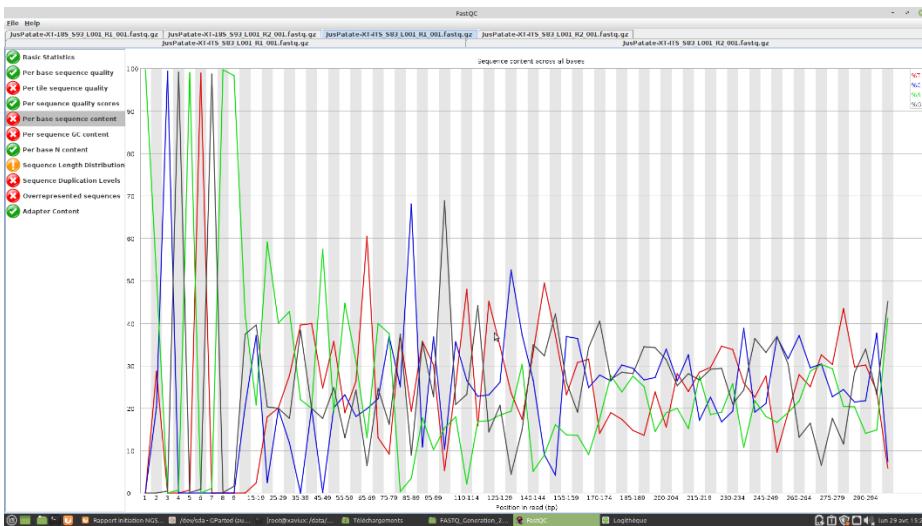


Figure 5 Résultat de l'analyse qualité par base de l'ITS du jus de patate en forward

Reverse:

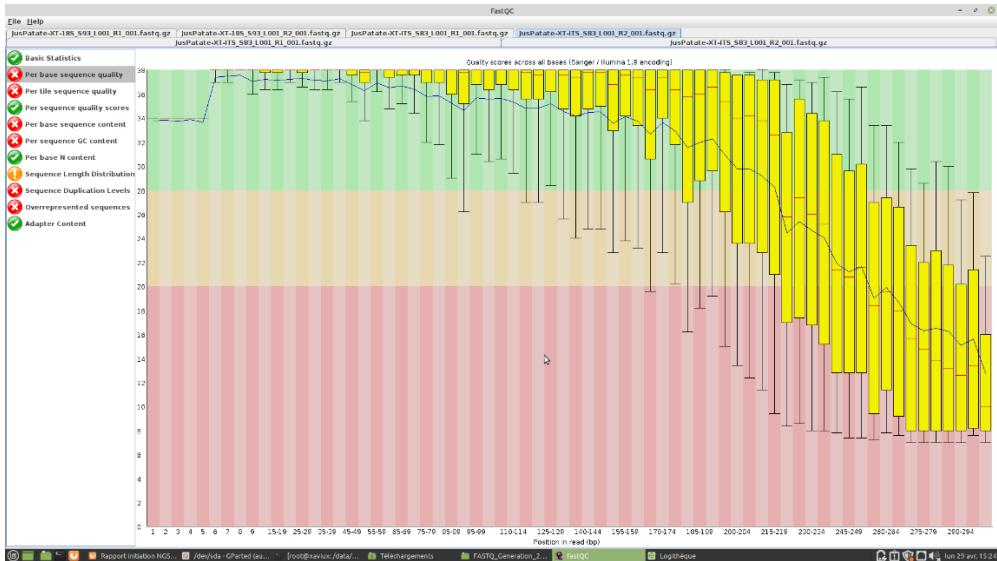


Figure 6 Résultats de l'analyse qualité par séquence de l'ITS du jus de patate en reverse

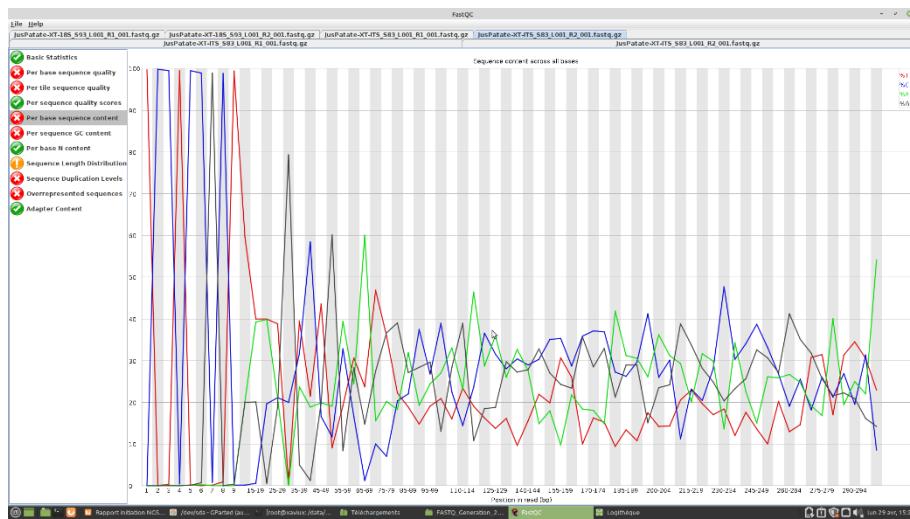


Figure 7 Résultats de l'analyse qualité par base de l'ITS du jus de patate en reverse

Il est observable des présences d'artefact d'adapteurs sur les sept premières bases et une qualité qui a tendance à chuter fortement en fin de séquence surtout sur les séquences en reverse.

Tous les éléments générés avec *FastQC* sont représentés en annexe 7.1.

4.1.2 Trimming

Au vu des analyses qualité présentées en Figure 4, Figure 5, Figure 6 et Figure 7 et en annexe 7.1, il est nécessaire de couper les parties possédant un score de mauvaise qualité et les artefacts issus des adaptateurs. Ainsi à l'aide du programme de *trimming Cutadapt*, les sept premières bases correspondant aux adaptateurs ont été coupées ainsi que les bases en 5' possédant un score de qualité inférieur à 20.

4.1.3 Taxonomie

L'analyse taxonomique a été réalisée à l'aide du programme Kraken2. Différentes bases de données ont été utilisées en fonctions du type de données. Ainsi, la base donnée *Silva* fut utilisée pour classifier les éléments 18S, la base donnée *Silva 138 16S* fut utilisée pour classifier les éléments 16S et la base de données *Fungi genome* de *RefSeq* pour les ITS.

	Eau	Terre	Compost	Jus de patate
16S	90,36%(125853/139279)	92,28% (159530/172870)	87,28% (3534/4049)	84,52% (4249/5027)
18S	94,95%(14812/15599)	89,61%(115936/129380)	91,07% (101825 /111804)	95,88% (153669/160270)
ITS	25,9%(14926/57624)	71,46%(124466/174171)	49,50% (250/505)	75,11% (172752/229988)

Tableau 3 Pourcentage de séquences reconnues dans chaque échantillon par Kraken2

Les résultats des ITS reconnus du compost sont faibles (49,50%), mais le nombre d'ITS acquis est également faible.

Le programme offre la possibilité de générer des fichiers permettant par la suite une visualisation interactive avec *Krona*.

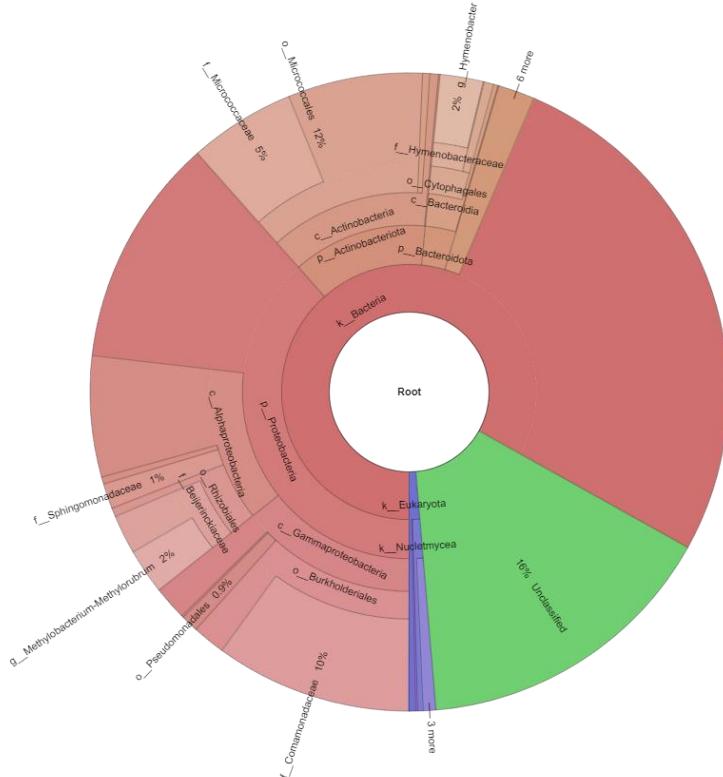


Figure 8 Représentation Krona du 16S Jus patate

La représentation de *Krona* montre tous les niveaux taxonomiques sur base des bases de données fournies pour l'analyse du règne jusqu'au genre, mais aussi montre leur abondance relative.

Ce graphique manipulable sur application web permet de focaliser la représentation taxonomique à certains embranchements.

Les différentes représentations générées pour chaque analyse sont présentées en annexe 7.2.

Ensuite il est possible de visualiser ces données sous une autre disposition à l'aide de *Pavian*.

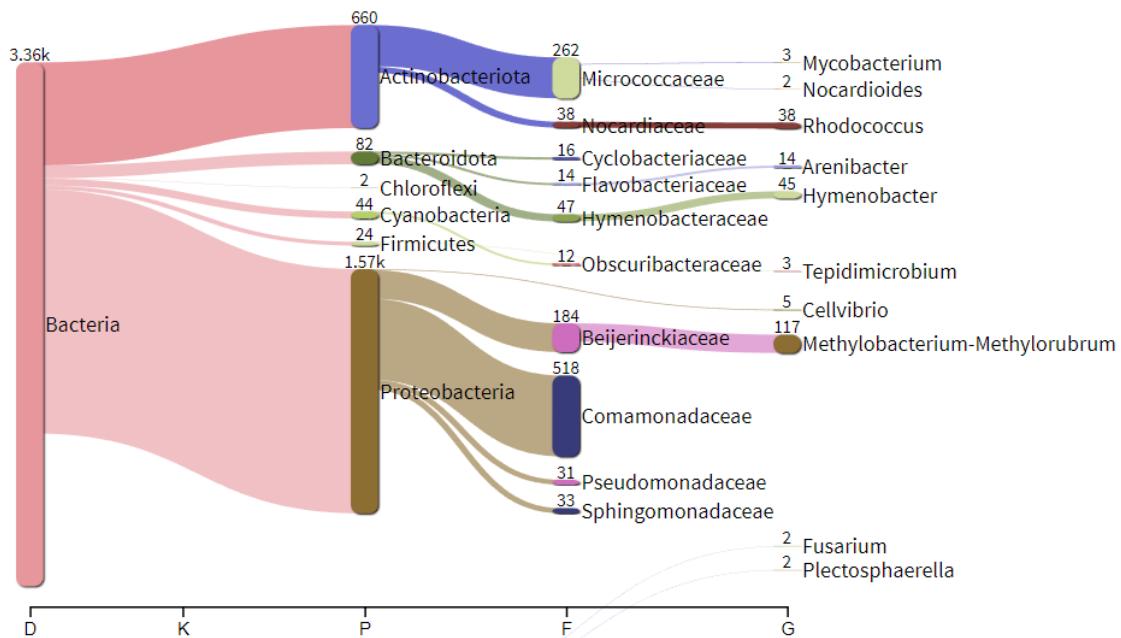


Figure 9 Représentation Pavian des 16S du compost

La représentation *Pavian* permet une visualisation sous forme d'arbre taxonomique, en mettant en valeur les abondances relatives des organismes retrouvées.

4.2 Analyse avec les données réalisées avec les données du pipeline

4.2.1 Analyse des phylums

L'analyse des résultats avec *Past* permet de générer des *Stacked Chart* pour visualiser la diversité globale des phylums présents dans les échantillons.

4.2.1.1 16S

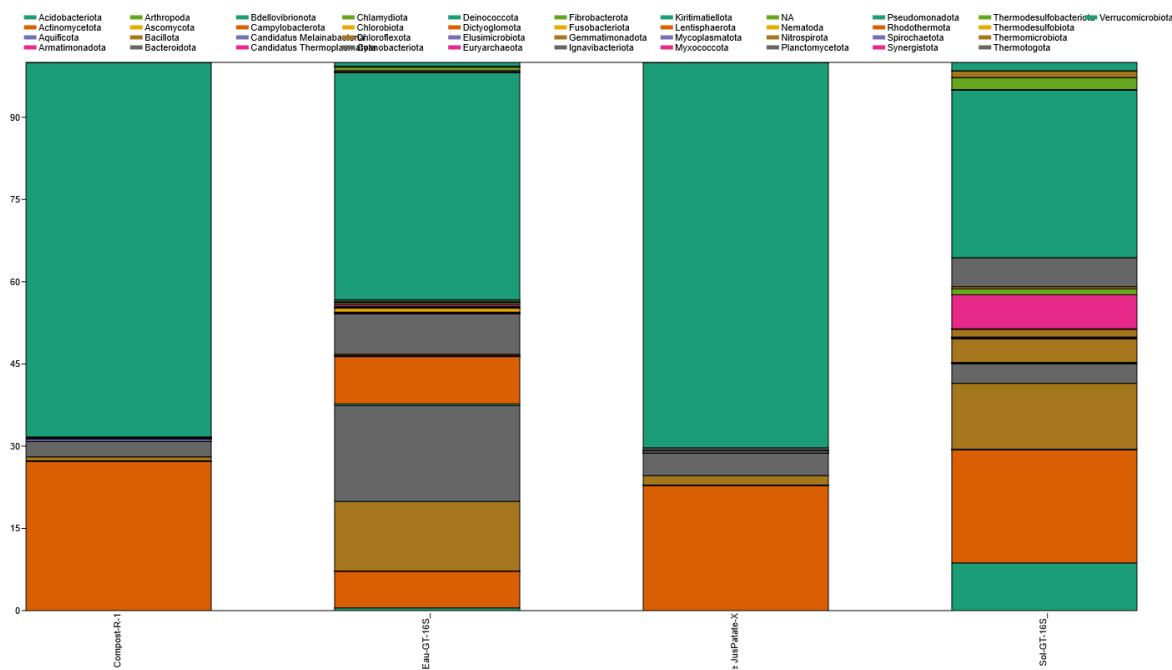


Figure 10 : Staked Charts des proportions des différents phylums dans les échantillons 16S

4.2.1.2 18S

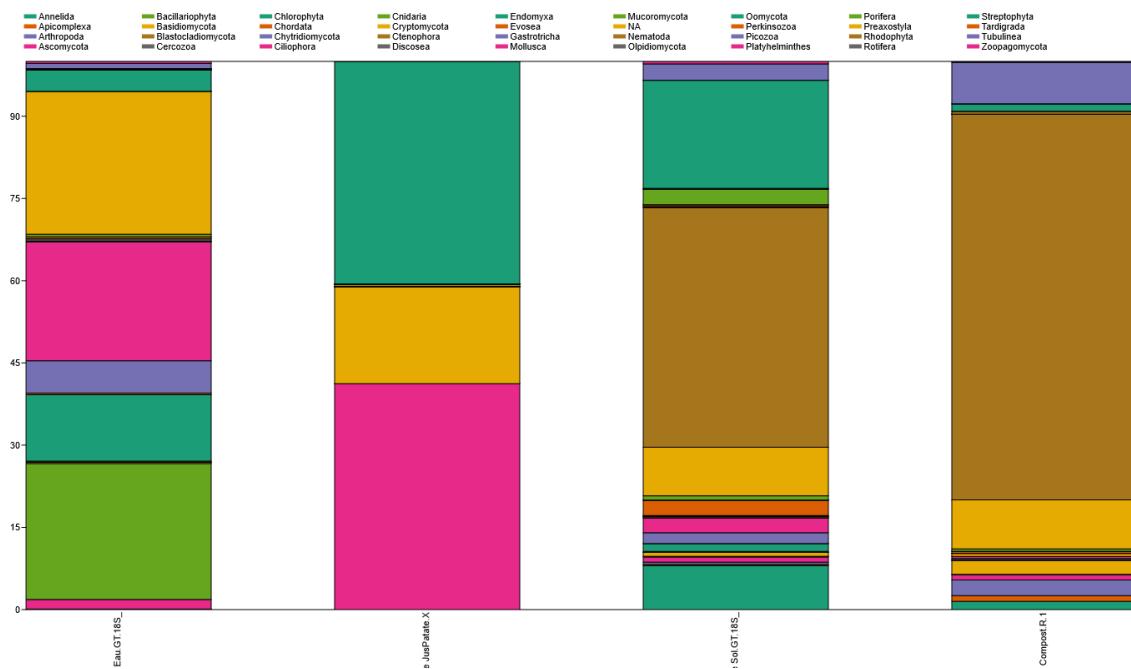


Figure 11 Staked Charts des proportions des différents phylums dans les échantillons 18S

L'analyse du compost met en valeur une présence prédominante de nématode dans l'échantillon.

4.2.1.3 ITS

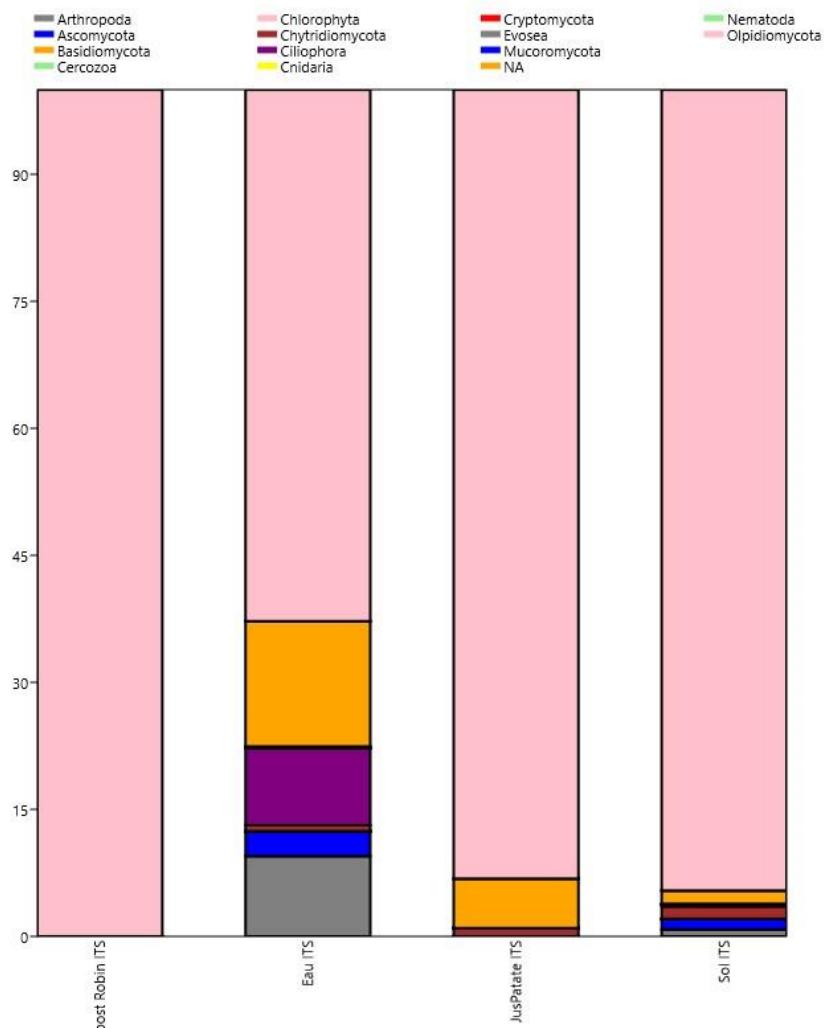


Figure 12 : Staked Chart des proportions des différents phylums dans les échantillons ITS

L'analyse du compost à ressorti seulement 148 phylums toute des Ascomycètes.

En analysant les résultants, il est apercevable que des organismes non considérés comme champignon sont compris dans les résultats. Voulant connaître la répartition de la population de champignon, les données ont été filtrées afin de garder que les organismes considérés comme champignon. Les organismes fongiques sont ainsi représentés si dessous

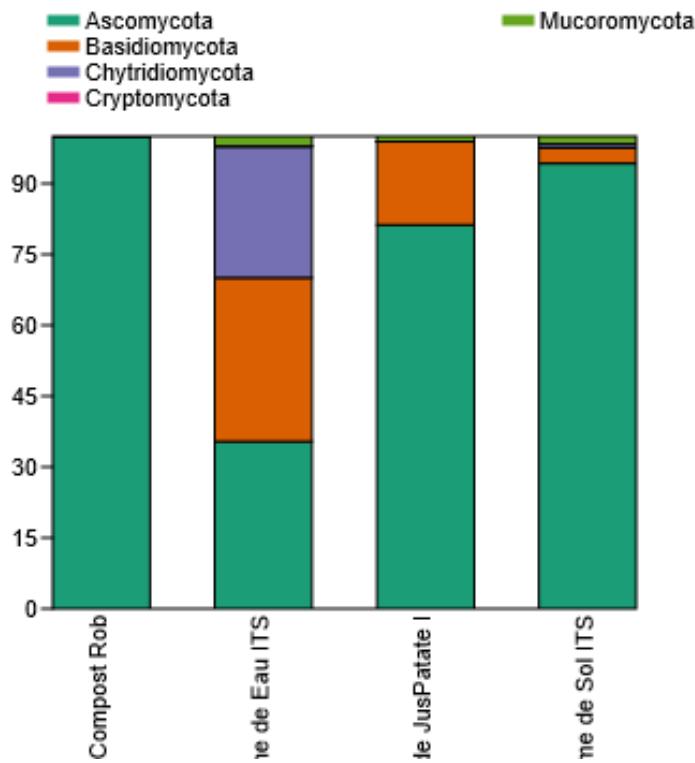


Figure 13 Staked Chart proportion des différents phylums de champignon dans les échantillons ITS

4.2.2 Analyse des genres

Afin de mettre en valeur les organismes présents dans les échantillons, les 10 organismes les plus présents ont été mis en valeur.

4.2.2.1 Compost

4.2.2.1.1 16S

Genres	Compost-R-16S
Arthrobacter	33%
Sphingomonas	32%
Variovorax	32%
Methylobacterium	12%
Stutzerimonas	5%
Afipia	3%
Hymenobacter	2%
Roseateles	2%
Bradyrhizobium	2%
Autre	10%

Tableau 4 Tableau des genres les plus présents dans la population 16S du compost

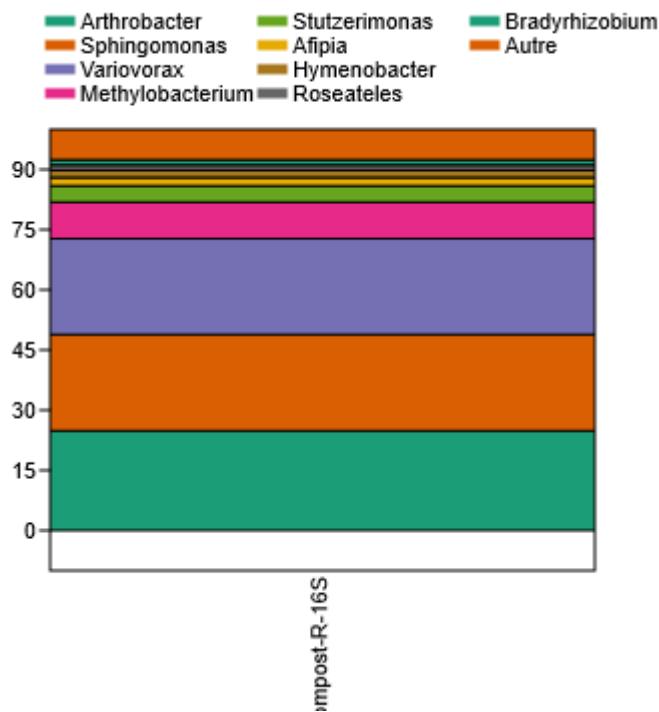


Figure 14 Stacked Chart des genres les plus présents dans la population 16S du compost

4.2.2.1.2 18S

Genres	Proportion
Pelodera	46%
NA	8%
Nolandella	7%
Mesorhabditis	7%
Neodiplogaster	4%
Sorodiplophrys	4%
Rhabditella	2%
Ceratellopsis	2%
Hypogastrura	2%
Autre	18%

Tableau 5 Tableau des genres les plus présents dans la population 18S du compost

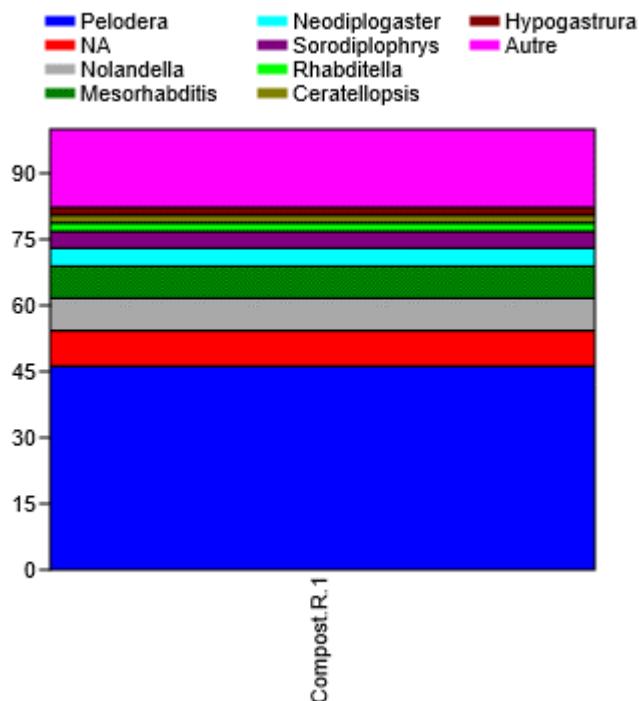


Figure 15 Stacked Chart des genres les plus présents dans la population 16S du compost

4.2.2.1.3 ITS

Étiquettes de lignes	Compost Robin ITS
Rinaldiella	41%
Aplosordaria	23%
Fusarium	22%
Podospora	8%
Cercophora	6%

Tableau 6 Tableau des genres les plus présents dans la population ITS du compost

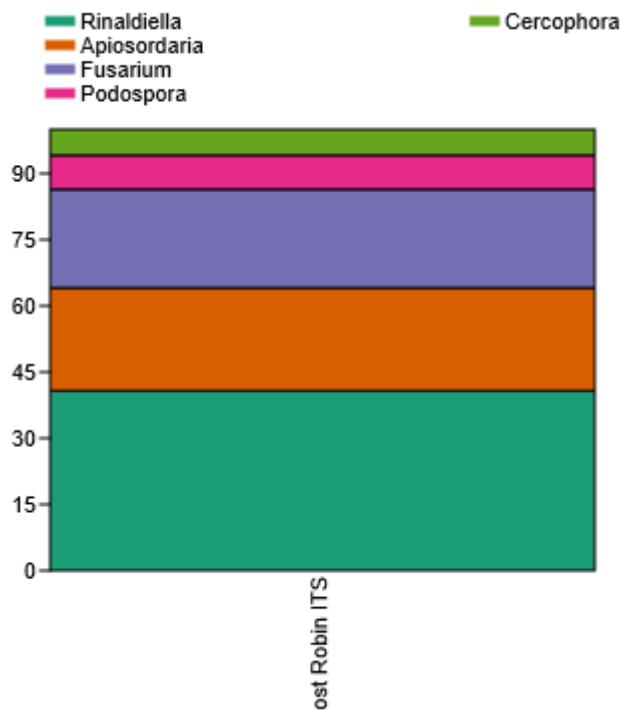


Figure 16 Stacked chart des genres les plus présents dans la population 16S du compost

Comme vu dans les analyses des phylums, les résultats d'ITS sont peu nombreux et peu diversifier.

4.2.2.2 Eau

4.2.2.2.1 16S

Tableau 7 : Top 10 16S Eau

Genres	Proportion (%)
Limnohabitans	9,35799878
Flavobacterium	8,45622321
NA	6,00892154
Aliarcobacter	5,85847984
Pleurocapsa	4,76865215
Acinetobacter	3,90011371
Rhodoferax	3,89399108
Arcobacter	1,96886207
Aquirufa	1,8848946
Hydrogenophaga	1,71346103

4.2.2.2.2 18S

Genres	Proportion (%)
Stephanodiscus	12,3971876
NA	11,4154948
Paraphysomonas	6,53356328
Sorodiplophrys	5,45900769
Zoothamnopsis	5,43910852
Mallomonas	4,94162908
Carteria	4,37781905
Navicula	3,86044044
Microglena	3,362961
Conticribra	3,17060228

Tableau 8 : Top 10 genre eau 18S

4.2.2.2.3 ITS

Genres	Proportion (%)
Streptophyta	23,5651822
Ascomycota	19,1945914
Basidiomycota	18,8124633
Chytridiomycota	15,0848765
NA	14,5502646
Ciliophora	4,67188419
Chlorophyta	2,69878013
Mucoromycota	1,12617578
Rhodophyta	0,2590388
Cryptomycota	0,03674309

Tableau 9 : Top 10 genre eau ITS

4.2.2.3 Terre

4.2.2.3.1 16S

Genres	Proportion (%)
NA	13,8155799
Vicinamibacter	4,66758709
Streptomyces	2,14528462
Acidibacter	1,84526338
Actinomarinicola	1,42974724
Rhodoplanes	1,4231096
Paludibaculum	1,32354503
Phaselicystis	1,28106415
Nocardioides	1,2306181
Dehalogenimonas	1,20141249

Tableau 10 : Top 10 genre sol 16S

4.2.2.3.2 18S

Genres	Proportion (%)
NA	21,1629804
Spondias	19,4365814
Neodiplogaster	13,1482992
Prismatolaimus	9,53637512
Eisenia	2,84667914
Amplimerlinius	2,13740824
Nolandella	2,06384238
Ceratiomyxella	1,94069952
Fabomonas	1,82315406
Plakina	1,77037854

Tableau 11: Top 10 genre sol 18S

4.2.2.3.3 ITS

Genres	Proportion (%)
Ascomycota	90,7912164
Basidiomycota	3,07289539
Streptophyta	1,64011537
Mucoromycota	1,55965688
Ciliophora	1,31147338
Chytridiomycota	0,86090584
Chlorophyta	0,42271653
NA	0,2432322
Rhodophyta	0,08169631
Cnidaria	0,00495129

Tableau 12 : top 10 genre sol ITS

4.2.2.4 Jus de patate

4.2.2.4.1 16S

Étiquettes de lignes	Jus de patate ITS
Sphingomonas	1223
Variovorax	902
Arthrobacter	896
Methylobacterium	377
Hymenobacter	141
Moraxella	90
Afipia	86
Stutzerimonas	61
Roseateles	55
Acidovorax	43
NA	33

Figure 17 Tableau des genres les plus présents dans la population 16S du Jus de patate

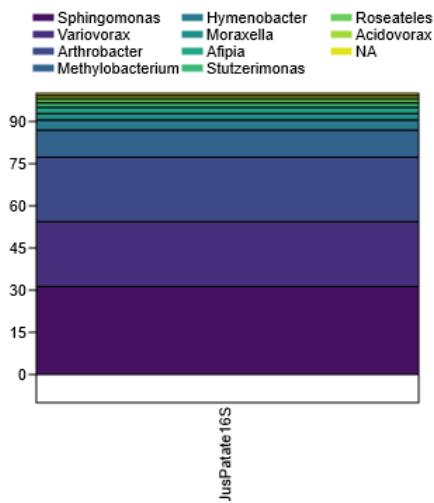


Figure 18 Stacked chart des genres les plus présents dans la population 16S du Jus de patate

4.2.2.4.2 18S

Étiquettes de lignes	Jus de patate ITS
NA	69153
Apotrichum	22259
Geotrichum	18104
Kurtzmaniella	10729
Pichia	3363
Nakazawaea	2660
Debaryomyces	1470
Cyberlindnera	1242
Yamadazyma	951

Meyerozyma	849
Wickerhamomyces	827

Figure 19 Tableau des genres les plus présents dans la population 18S du Jus de patate

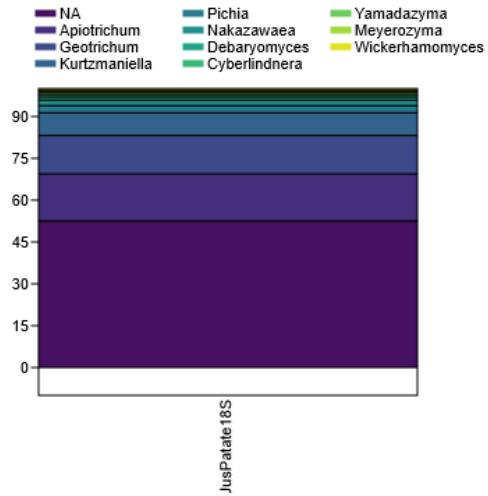


Figure 20 Stacked chart des genres les plus présents dans la population 16S du Jus de patate

4.2.2.4.3 ITS

Étiquettes de lignes	Jus de patate ITS
Penicillium	62599
Apotrichum	30163
Geotrichum	22690
Cladosporium	22540
Ipomoea	11992
Botrytis	11199
Spathaspora	11173
NA	7749
Kurtzmaniella	4407
Rhizopus	1237
Talaromyces	1101

Figure 21 Tableau des genres les plus présents dans la population ITS du Jus de patate

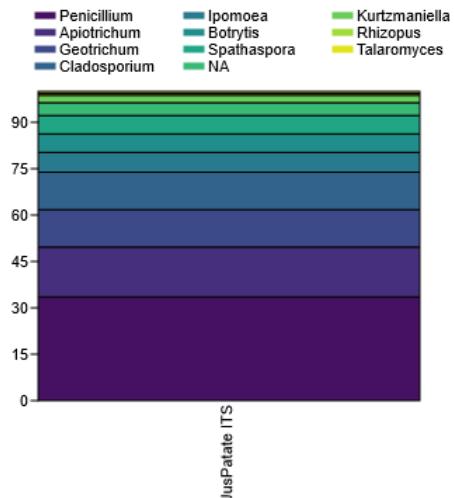


Figure 22 Stacked chart des genres les plus présents dans la population 16S du Jus de patate

4.2.3 Calcul de l'indice de diversité

L'indice de diversité est une mesure utilisée pour évaluer la diversité des espèces dans un échantillon. Il existe plusieurs indices de diversité, mais les plus courants sont l'indice de Shannon et l'indice de Simpson, voici les calculs de ces indices sur les données issues du pipeline d'Ath :

4.2.3.1 Indice de Shannon

L'indice se calcule à l'aide de la formule suivante :

$$H = - \sum (p_i * \log(p_i))$$

Où p_i est la proportion de l'espèce i par rapport au total des espèces.

Voici les différents indices de diversité de Shanno n pour les données ITS,16S,18S calculés avec la formule ci-dessus :

Tableau 13 : Indice de diversité ITS sur les phylums

Avec Pipeline d'Ath	Composte ITS	Eau ITS	Jus de patate ITS	Sol ITS
Indice de diversité de Shannon	0	0,8	0,31	0,20

Avec Pipeline d'Ath et Past	Composte ITS	Eau ITS	Jus de patate ITS	Sol ITS
Indice de diversité de Shannon	0	1,84	0,7	0,45

Avec Notre Pipeline et past	Composte ITS	Eau ITS	Jus de patate ITS	Sol ITS
Indice de diversité de Shannon	0	0,06	0,0001	0

Le compost à un indice de 0 indique que cet échantillon est composé presque entièrement d'une seule espèce (ou groupe taxonomique), montrant ainsi une diversité extrêmement faible ou inexisteante.

L'eau présente un indice de 0,8 indiquant une diversité relativement élevée par rapport aux autres échantillons.

Jus de patate à un indice de 0,31 montre une diversité modérée. Cela pourrait indiquer une certaine dominance de quelques taxons avec la présence de divers autres groupes en plus petites quantités.

Le sol montre l'indice le plus bas parmi les échantillons ayant une diversité détectable, avec une valeur de 0,20.

Tableau 14 : Indice de diversité 18S sur les phylums

Avec Pipeline d'Ath	Composte 18S	Eau 18S	Jus de patate 18S	Sol 18S
Indice de diversité de Shannon	0,53	0,80	0,46	0,83

Avec Pipeline d'Ath et Past	Composte 18S	Eau 18S	Jus de patate 18S	Sol 18S
Indice de diversité de Shannon	1,02	1,7	1,046	1,772

Avec Notre Pipeline et past	Composte 18S	Eau 18S	Jus de patate 18S	Sol 18S
Indice de diversité de Shannon	1,162	1,9	0,67	1,4

Tableau 15 : Indice de diversité 16S sur les phylums

Avec Pipeline d'Ath	Composte 16S	Eau 16S	Jus de patate 16S	Sol 16S
Indice de diversité de Shannon	0,35	0,80	0,37	0,93

Avec Pipeline de Ath et Past	Composte 16S	Eau 16S	Jus de patate 16S	Sol 16S
Indice de diversité de Shannon	0,8	1,8	0,84	2,13

Avec Notre Pipeline et past	Composte 16S	Eau 16S	Jus de patate 16S	Sol 16S
Indice de diversité de Shannon	1,017	1,94	0,9958	2,3

Voici une remarque générale par rapport aux différents indices de diversité calculés. Nous les avons faites manuellement en suivant la formule indiquée plus haut et nous nous rendons compte qu'avec

les mêmes données les indices sont différents quand elles sont calculées avec *Past*. Ne sachant pas si *Past* utilise bien la formule décrite nous avons préféré mettre tous les résultats calculés.

Source : <https://drdk.me/metagenomique.html>

4.3 Comparaison des résultats

Les résultats générés par Kraken indiquent que les bases de données que le programme utilise des noms scientifiques anciens malgré le fait que nous avons sélectionné les plus récentes disponibles (2021).

4.3.1 Comparaison Compost

4.3.1.1 Phylum Compost 16S

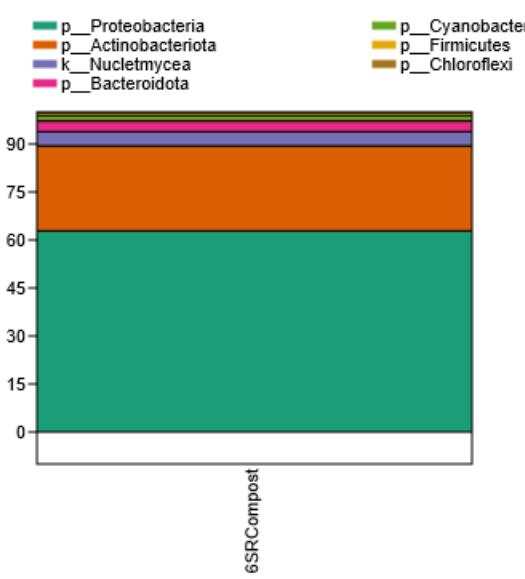


Figure 23 Stacked Chart des phylums des données analysées avec Kraken2

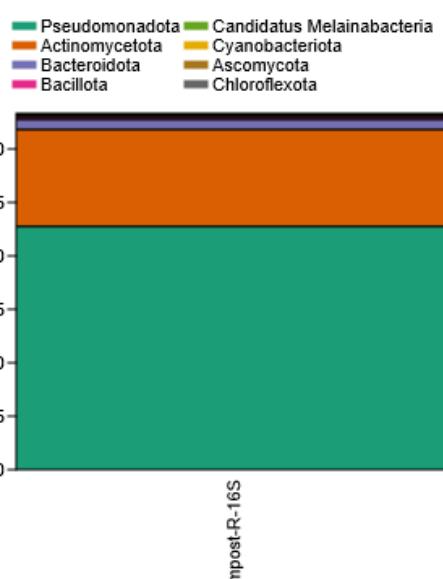


Figure 24 Stacked Chart des phylums des données analysées avec le pipeline d'Ath

En comparant les données, il est constatable que les abondances relatives sont à peu près les mêmes au niveau du phylum avec une abondance supérieure des *Pseudomonata* (*Proteobacteria*) à plus de 60% dans les deux échantillons.

L'analyse avec le pipeline d'Ath à identifier 3444 phylums contre 2488 phylums sur Kraken2.

Il est à noter que les ascomycètes ne sont pas retrouvés dans l'analyse avec Kraken2, cela provient du fait que l'analyse en 16S a été réalisée sur une base de données de bactéries donc ne pouvant pas ressortir d'ascomycète.

4.3.1.2 Phylum Compost 18S

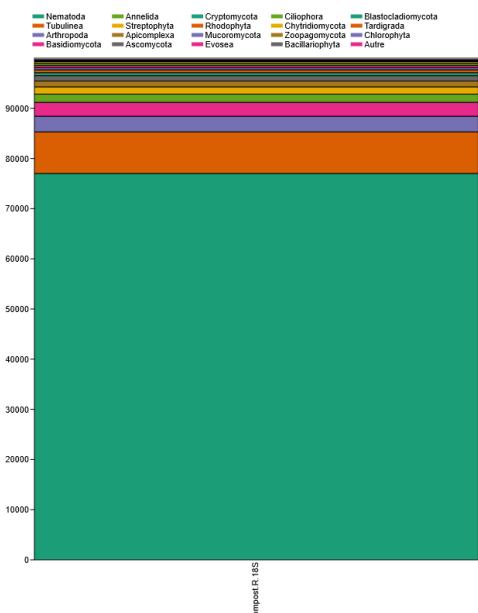


Figure 25 Stacked Chart des phylums des données analysées avec le pipeline d'Ath

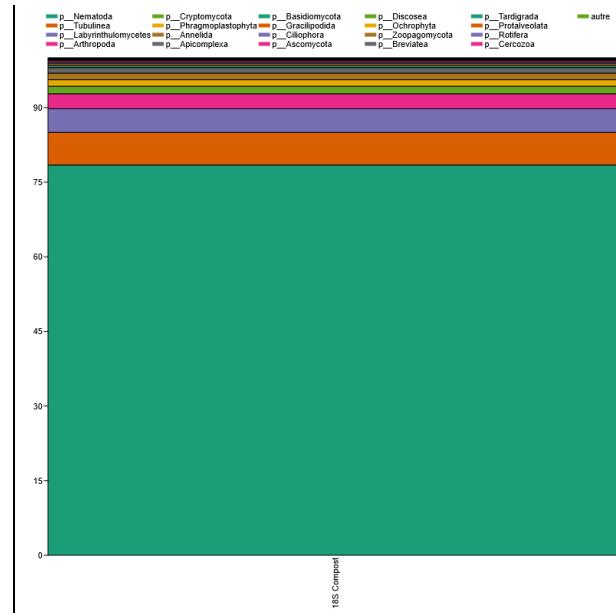


Figure 26 Stacked Chart des phylums des données analysées avec Kraken2

En comparant les données, il est constatable que les abondances relatives sont à peu près les mêmes au niveau du phylum avec une abondance supérieur des Nématodes à plus de 75% dans les deux échantillons.

4.3.1.3 Phylum Compost ITS

Les deux analyses ont trouvé seulement des *Ascomycètes*, le pipeline d'Ath dénombre 148 ITS trouvées et Kraken2 en dénombre 128.

4.3.2 Comparaison Eau

Phylum Eau 16S

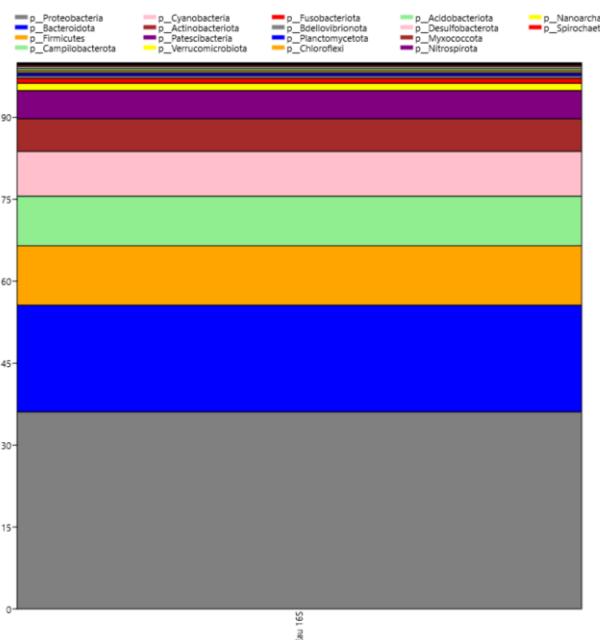


Figure 27 : Stacked Chart des phylums Eau 16S des données analysées avec notre pipeline



Figure 28 : Stacked Chart des phylums Eau 16S des données analysées avec le pipeline d'Ath

Proteobacteria : Il y a une augmentation significative de ce phylum dans l'échantillon traité avec le pipeline d'Ath. 47417 comparé à 31707 dans l'échantillon traité avec notre façon de faire.

Bacteroidota : Une légère augmentation est observée dans Eau 16S Ath avec 19959 comparé à 17141 avec nos données.

Nous retrouvons les phylums principaux en comparaison avec le pipeline d'Ath, mais nos quantités sont moindres et parfois significativement inférieures aux résultats d'Ath.

4.3.2.1 Phylum Eau 18S

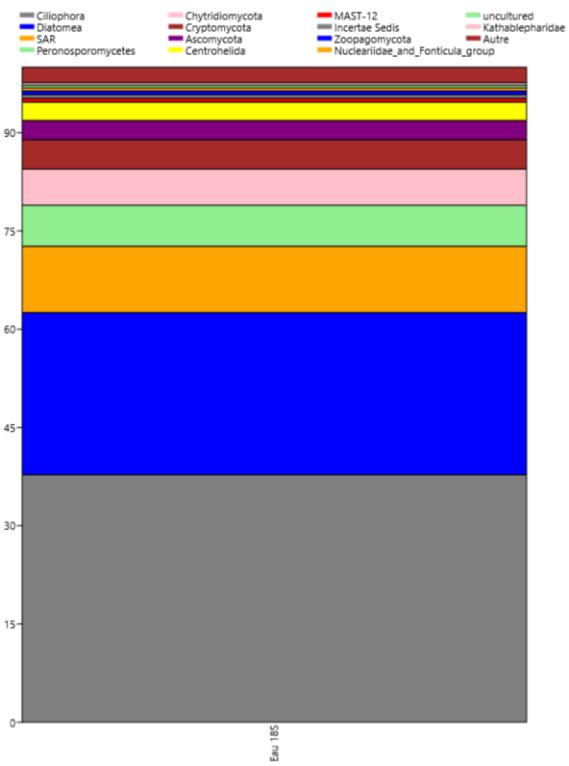


Figure 29 : Stacked Chart des phylums Eau 18S des données analysées avec notre pipeline

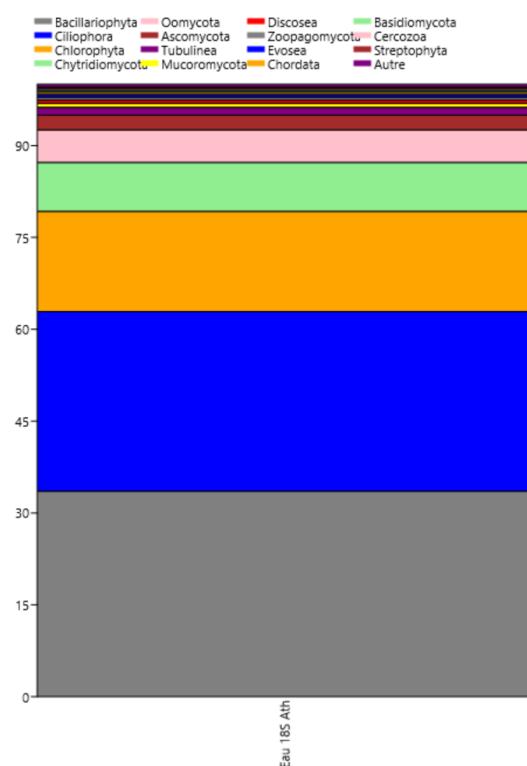


Figure 30 : Stacked Chart des phylums des données Eau 18S analysés avec le pipeline d'Ath

4.3.2.2 Phylum Eau ITS

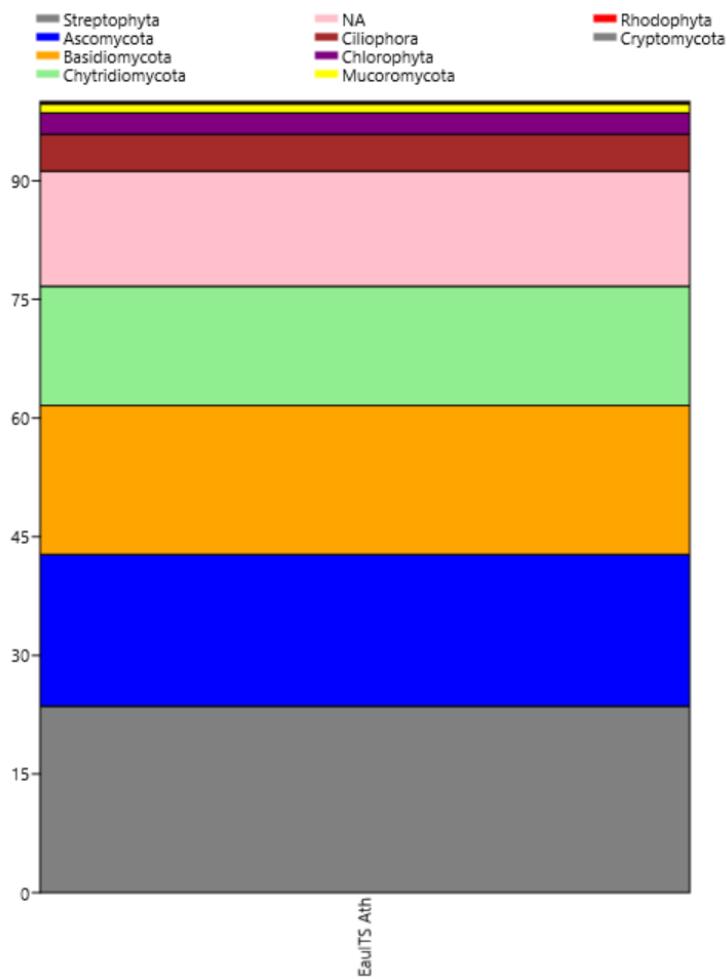


Figure 31 : Stacked Chart des phylums des données Eau ITS analysées avec le pipeline d'Ath

L'analyse réalisée avec notre pipeline ne révèle que le phylum des *Ascomycota* et des *Basidiomycota* pour l'échantillon d'eau ITS.

4.3.3 Comparaison Sol

4.3.3.1 Phylum Sol 16S

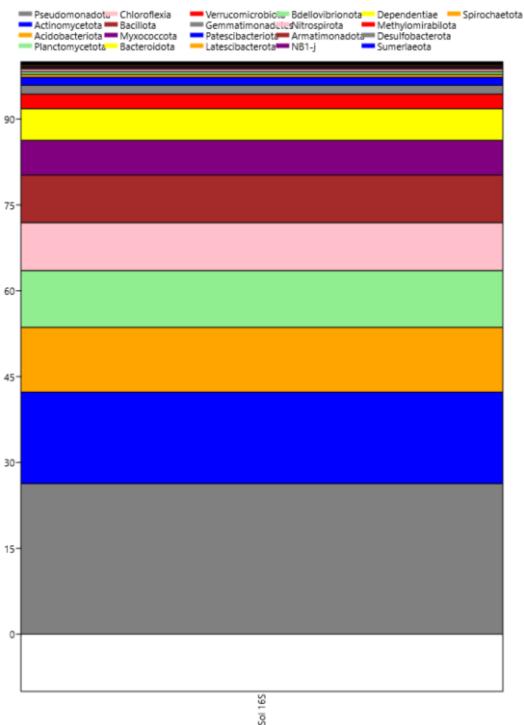


Figure 32 : Stacked Chart des phylums Sol 16S des données analysées avec notre pipeline

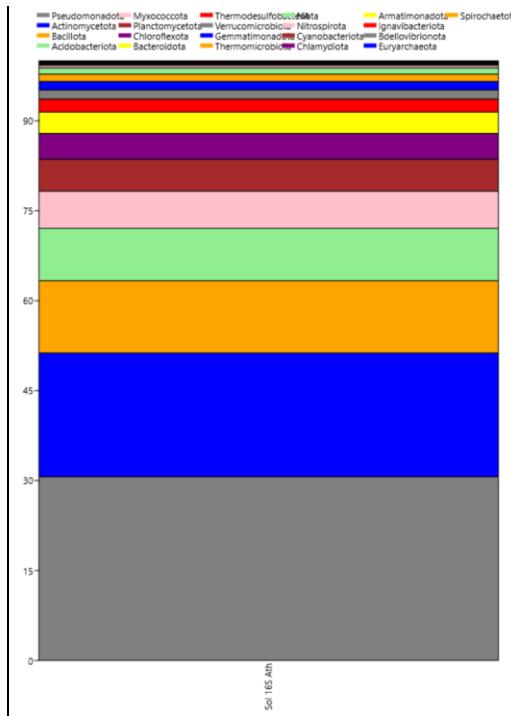


Figure 33 : Stacked Chart des phylums Sol 16S des données analysées avec le pipeline d'Ath

Il est observable une diminution générale des quantités de phylums analysés avec notre technique.

4.3.3.2 Phylum Sol 18S

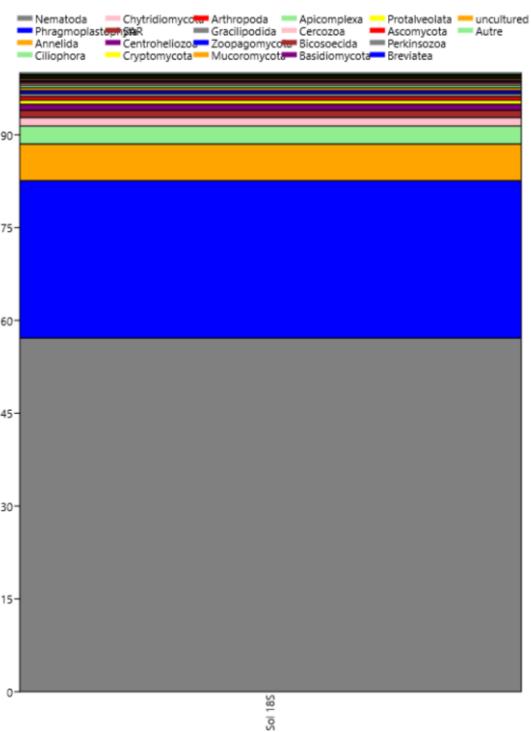


Figure 34 : Stacked Chart des phylums Sol 18S des données analysées avec notre pipeline

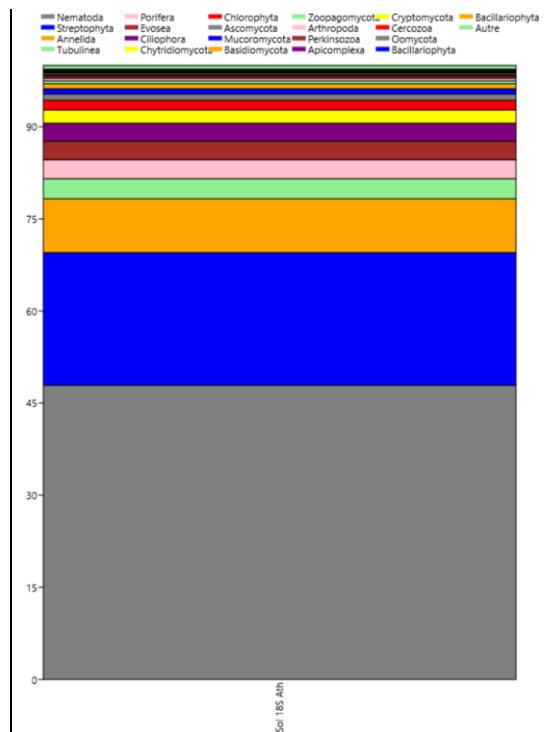


Figure 35 : Stacked Chart des phylums Sol 18S des données analysées avec le pipeline d'Ath

4.3.3.3 Phylum Sol ITS

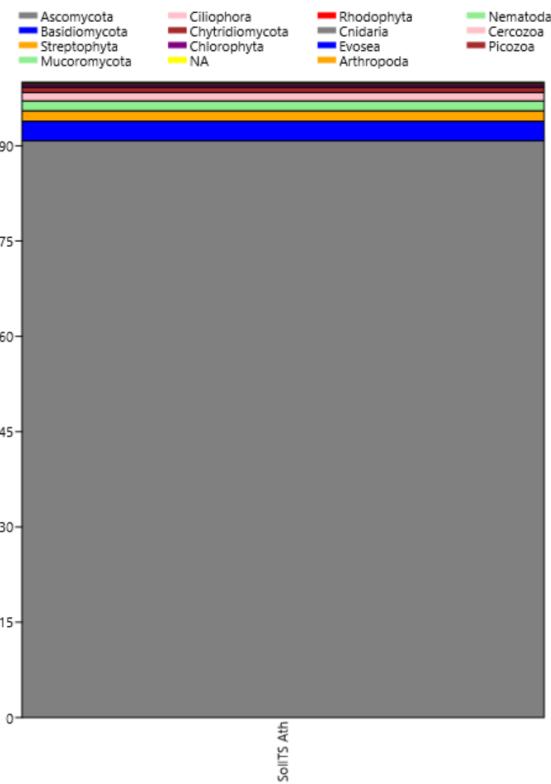


Figure 36 : Stacked Chart des phylums Sol ITS des données analysées avec le pipeline d'Ath

L'analyse réalisée avec notre pipeline ne révèle que le phylum des *Ascomycota* pour l'échantillon de sol ITS.

4.3.4 Comparaison Jus de patate

4.3.4.1 Phylum patate 16S

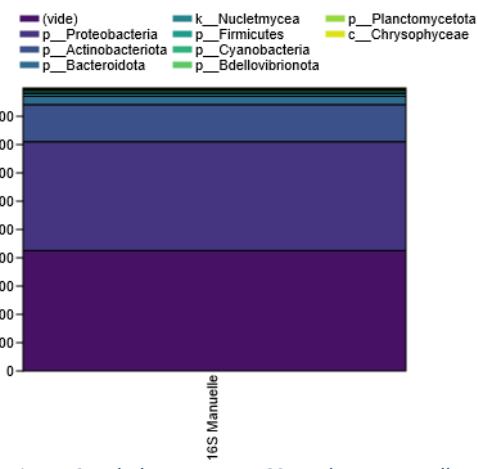


Figure 37 Phylum patate 16S analyse Manuelle

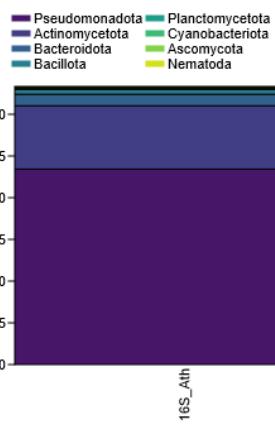


Figure 38 Phylum patate 16S analyse Ath

Outre les *protéobactéries* et les *pseudomonadota* étant le même phyla, les résultats sont plutôt similaires.

4.3.4.2 Phylum patate 18S

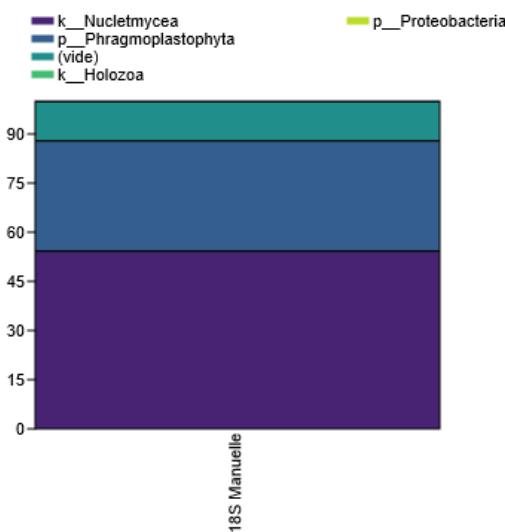


Figure 39 Phylum patate 18S analyse manuelle

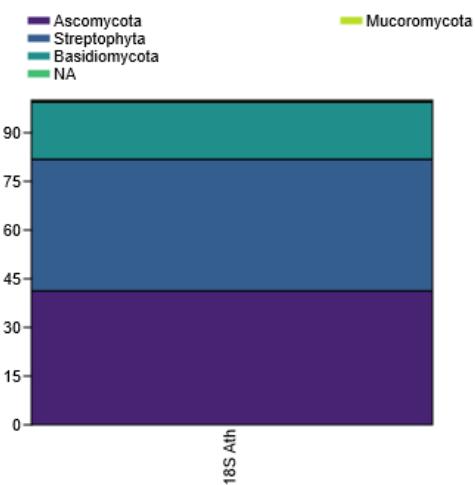


Figure 40 Phylum patate 18S analyse Ath

Ici, on peut remarquer une différence entre les individus trouvés mais une quantité d'inconnus plutôt similaire.

4.3.4.3 Phylum patate ITS

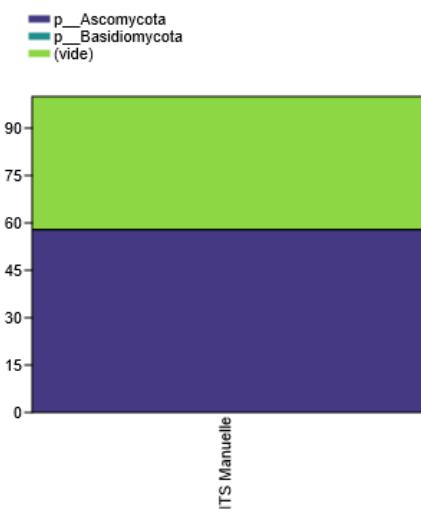


Figure 41 Phylum patate ITS analyse manuelle

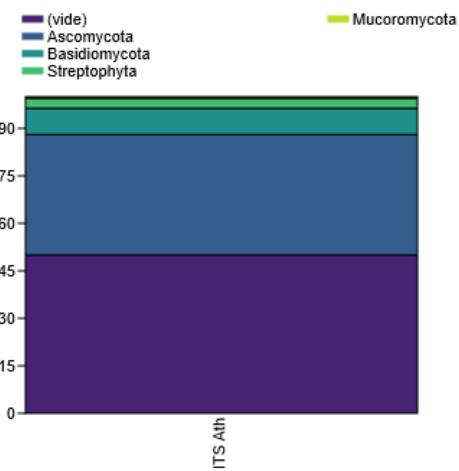


Figure 42 Phylum patate ITS analyse Ath

Les analyses ITS donnent des résultats plutôt similaires avec une quantité d'inconnus très élevé.

4.4 Interprétation des résultats

4.4.1 Compost

Les résultats présentés en partie 4.2.2.1.1 démontre que dans la population 16S il y a une forte présence de bactérie du genre des *Arthrobacter*. Ces organismes, souvent retrouvés dans le sol, interviennent dans la dégradation d'hydrocarbures, de tannins, la cellulose, de lignine et certains herbicide et pesticides. De plus, la bactérie possède une activité de promotion de la croissance. En effet, la bactérie peut jouer un rôle important dans la protection contre le stress abiotique lorsqu'une plante grandit dans les milieux pollué (2). La présence de ces bactéries contribue alors à l'efficacité du compost sur les cultures.

L'analyse de la population de 18S présentés en partie 4.2.2.1.2 révèle une présence importante de nématode du genre *Pelodera* en proportion importante (46%). Ces vers non-segmentés sont saprophytes, ce qui signifie qu'elles se nourrissent principalement de matière organique morte et en décomposition. Elle joue donc un rôle essentiel dans les processus de décomposition et au recyclage des éléments nutritifs où elle vit. La présence de ces nématodes est donc essentielle au processus de transformation des déchets organiques. (3)

L'analyse des ITS montre des résultats anormalement bas et donne peu d'espace à une utilisation de ces données.

4.4.2 Eau

Dans la population 16S de l'eau, nous retrouvons une population de plus ou moins 10% de bactéries venant du genre des *Flavobacterium*. Grams négatifs, elles appartiennent à la famille des *Flavobacteriaceae*. Elles sont généralement jaunes en raison de pigments caroténoïdes. Ces bactéries se retrouvent généralement dans l'eau douce ainsi que l'eau de mer et le sol. Elles jouent un rôle crucial dans la décomposition de la matière organique, facilitant ainsi le recyclage des nutriments dans leurs écosystèmes. Certaines espèces sont pathogènes pour les poissons causant des maladies comme la flavobactéries. (4)

En ce qui concerne les résultats de nos 18S, nous pouvons parler des *Stephanodiscaceae* qui est le genre le plus présent dans l'échantillon. C'est une famille de diatomées, un groupe de microalgues qui possèdent des parois cellulaires siliceuses distinctives, connues sous le nom de frustules. Elles sont principalement trouvées dans les environnements aquatiques (eaux douces) et sont souvent associées à des milieux eutrophiques, où les nutriments sont abondants. Une des utilités de cette famille est la suivante : les fossiles de *Stephanodiscaceae*, préservés dans les sédiments, sont utilisés pour reconstruire les conditions environnementales passées, offrant des aperçus sur les changements climatiques historiques. (5)

Au niveau de la population des ITS, il est retrouvable comme genre prédominant les *Plagiomnium*. C'est un genre de mousse appartenant à la famille des *Mniaceae* se retrouvant principalement dans les milieux humides et ombragés comme les sous-bois et les bords de ruisseaux. (6)

4.4.3 Terre

Pour commencer, nous allons parler des résultats des 16S. La famille que nous allons aborder est celle des *Streptomyces*. Même si cette dernière n'est pas prédominante dans notre échantillon c'est

intéressant de s'y attarder. C'est est un genre de bactéries Gram positif qui appartient à l'ordre des *Actinomycetales*. Elles sont célèbres pour leur rôle crucial dans la production naturelle d'antibiotiques. La famille *Streptomyces* est largement répandue dans les sols et contribue de manière significative à la décomposition de la matière organique. (7)

Pour faire suite, la famille la plus présente est celle des *spondias* (19,43%). C'est un genre de plantes appartenant à la famille des *Anacardiaceae*, qui comprend également les manguiers et les pistachiers. Ce genre est principalement connu pour ses arbres fruitiers tropicaux. (8)

Pour finir, les ITS. Le *Cephalotrichum* (13% dans l'échantillon) est un genre de champignons microscopiques appartenant à la famille des *Trichocomaceae*. Ces champignons sont souvent retrouvés sous forme de moisissures. Bien que principalement non pathogène pour les plantes, leur présence excessive peut affecter la qualité des sols et des récoltes par leur compétition pour les ressources. (9)

4.4.4 Jus de patate

L'analyse minutieuse de la communauté microbienne provenant d'un échantillon de patate douce moisie révèle une diversité impressionnante de levures et de bactéries qui cohabitent avec cette culture. Les résultats du séquençage 18S ont mis en avant une gamme de genres de levures, laissant entrevoir des interactions complexes dans l'écosystème de la patate douce. Bien que ces levures ne comprennent pas de pathogènes avérés de la patate douce, leur présence suggère une influence potentielle sur l'environnement racinaire et la santé des plantes.

De même, les données du 16S ont révélé une multitude de bactéries, chacune apportant sa contribution à la diversité microbiologique associée à la patate douce. Des groupes comme les *Pseudomonadota* et les *Actinomycetota* sont courants dans le sol et peuvent influencer divers aspects de la santé des plantes tels que la nutrition ou les maladies. Bien que nous ne constatons pas la présence de pathogène bien connue, leur rôle dans l'équilibre écologique reste crucial à comprendre. (10)

Cependant, les résultats ITS appellent à une attention particulière, avec la détection de genres de moisissures et de champignons tels que *Penicillium* et *Botrytis*, qui sont sûrement responsables de la pourriture de la patate douce. (11) (12)

5 Conclusion

Ce laboratoire a permis de réaliser une analyse métagénomique en partant de l'étape du prélèvement d'échantillon jusqu'au séquençage avec les technologies NGS puis au traitement des résultats et enfin à l'identification d'organisme d'intérêts présents dans les échantillons.

La phase de préparation des échantillons a révélé plusieurs biais potentiels pouvant altérer la qualité des résultats. Parmi ceux-ci, on note la présence de contaminations d'ARN dans tous les échantillons, ainsi que des erreurs d'insertion lors de la préparation des librairies. De plus, certaines étapes ont dû être répétées en raison d'erreurs de manipulation, soulignant l'importance de la précaution à chaque étape du processus.

Il a été réalisé l'analyse en partant des fichiers *FASTQ* générés à l'issue du séquençage. Il a été constaté en réalisant les analyses qualité que les séquences n'étaient pas de très bonne en reverse et que l'on a retrouvé la présence d'artefacts d'adaptateur. Il s'est donc montré nécessaire de réaliser un *trimming* pour enlever les bases de mauvaise qualité. Par la suite, l'analyse taxonomique fut réalisée à l'aide du programme *Kraken2* sur les bases de données *SILVA* et *fungi génome de refseq*. Une visualisation taxonomique des organismes et de leur proportion a ensuite été réalisée en utilisant les programmes *Krona* et *Pavian*.

Une analyse a été réalisée sur les données issues du pipeline d'Ath, les données ont été visualisées à l'aide de *Past*. Ainsi, les proportions des phylums présents dans les échantillons ont été mises en avant à l'aide de stacked chart généré sur *Past*. Par la suite, une seconde stacked chart, a mis en évidence les genres les plus présents permettant de s'intéresser aux organismes présents et de pouvoir en développer un d'intérêt.

Afin de rester critique sur les résultats générés avec *Kraken2*, une comparaison des résultats réalisés de cette méthode nouvellement testée avec ceux de la méthode du pipeline d'Ath fut réalisée. Il fut alors visible que les deux méthodes, bien qu'ayant effectué des analyses sur des bases de données différentes, obtiennent des résultats similaires bien que la proportion ne soit pas exactement similaire. Il aurait été intéressant de réaliser des tests de corrélation entre les résultats, mais qui se serait montré fastidieux à réaliser à cause de stratégie de labélisation réalisée par les bases de données fournies différentes. Un changement des paramètres de *Kraken2* sur les taux de confiances pour une classification pourrait peut-être faire coïncider plus les résultats.

Dans l'ensemble, cette analyse souligne l'importance cruciale de comprendre la diversité microbienne présente dans les différents échantillons que nous avons analysés. Une telle compréhension est essentielle pour développer des stratégies de gestion des microorganismes afin de gérer leurs impacts sur la qualité de l'eau ou même la gestion des environnements en général.

6 Bibliographie

1. QIAGEN. *DNeasy PowerSoil Kit*.
2. Pratiti Roy, Ashok Kumar. Arthrobacter. [En ligne] <https://doi.org/10.1016/B978-0-12-823414-3.00001-0>.
3. SCHULT, Franz. Description of Rhabditis (Pe1odera)pseudoteres n. sp. (Rhabditidae : Nematoda) with a redescription of its sibling R. (p.) teres (Schneider, 1866). [En ligne] 1989. https://horizon.documentation.ird.fr/exl-doc/pleins_textes/pleins_textes_5/pt5/nemato/27869.pdf.
4. Flavobacterium. [En ligne] <https://microbewiki.kenyon.edu/index.php/Flavobacterium>.
5. Stephanodiscaceae. [En ligne] <https://fr.wikipedia.org/wiki/Stephanodiscaceae>.
6. Plagiomnium. [En ligne] <https://fr.wikipedia.org/wiki/Plagiomnium>.
7. Streptomyces. [En ligne] <https://fr.wikipedia.org/wiki/Streptomyces>.
8. Spondias. [En ligne] <https://fr.wikipedia.org/wiki/Spondias>.
9. J.H.C. Woudenberg, 1 M. Sandoval-Denis,1,2 J. Houbraken,1,* K.A. Seifert,3 and R.A. Samson1. Cephalotrichum and related synnematosus fungi with notes on species from the built environment. [En ligne] 2017 Sep. PMID: 29158610.
10. Britannica, Encyclopedia. actinomycete. [En ligne] 16 Dec. 2022. <https://www.britannica.com/science/actinomycete>.
11. Penicillium spp. *in spq*. [En ligne] <https://www.inspq.qc.ca/moisissures/fiches/penicillium-spp>.
12. Nicholas Cheung, Lei Tian, Xueru Liu, and Xin Li. The Destructive Fungal Pathogen *Botrytis cinerea*—Insights from Genes Studied with Mutant Analysis. *ncbi*. [En ligne] 2020 Nov 7. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7695001/>.

7 Annexe

7.1 FastQC

7.1.1 FASTQC Compost

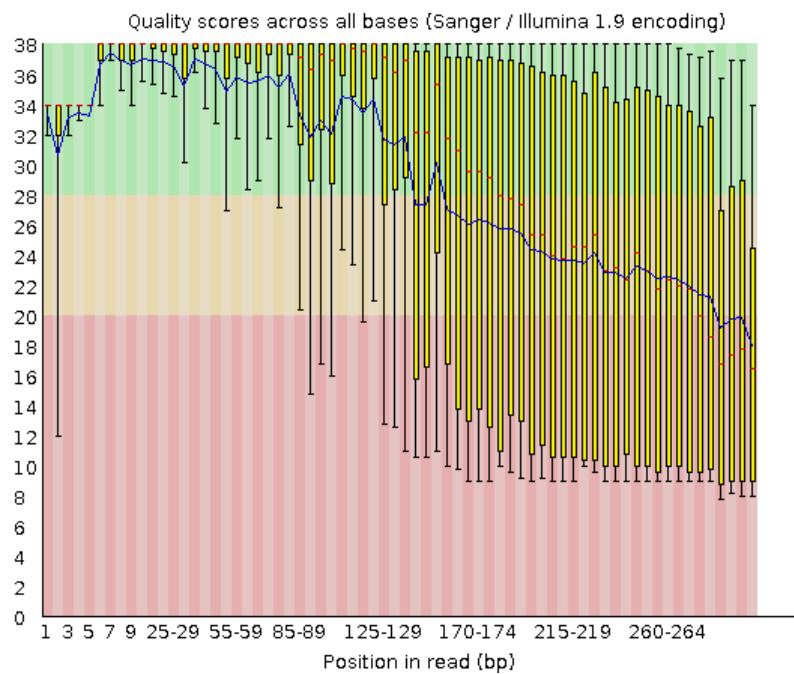


Figure 43 : FASTQC quality per base sequence quality R1 ITS Compost Robin

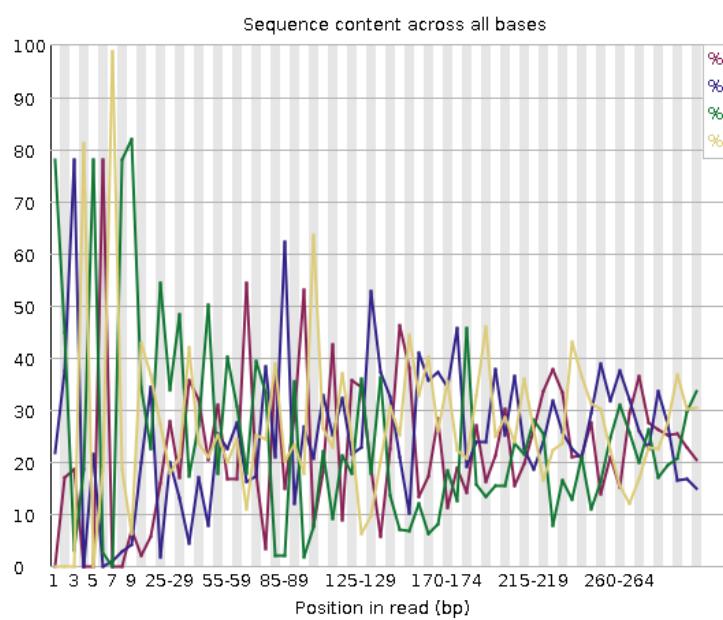


Figure 44 FASTQC quality per base sequence content R1 ITS Compost Robin

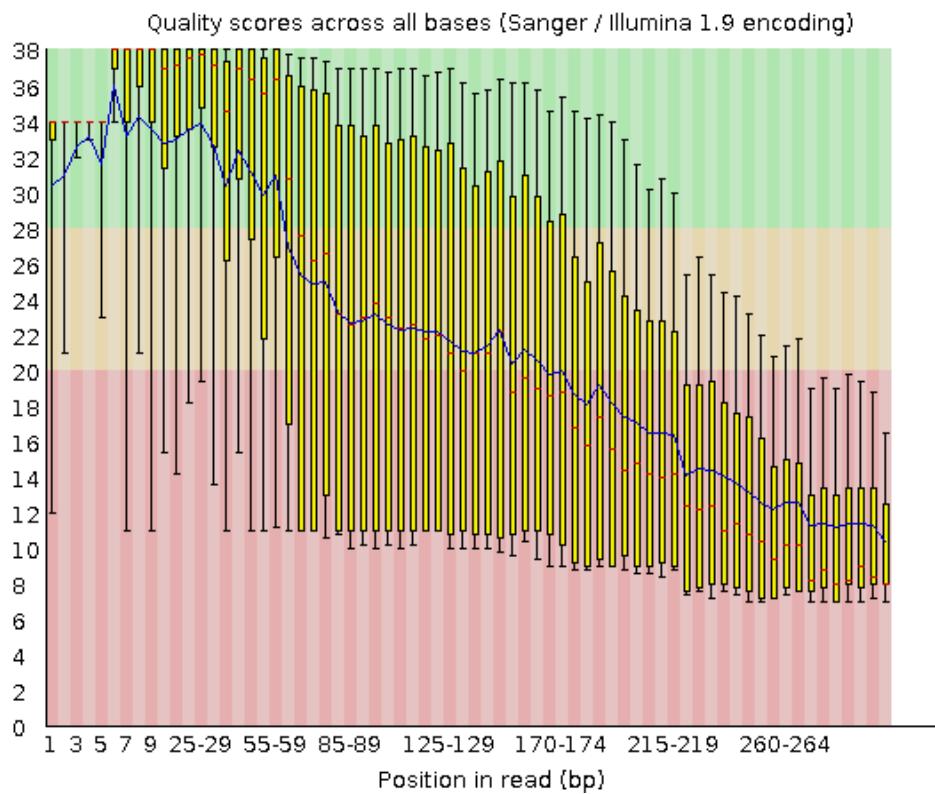


Figure 45 FASTQC R2 FASTQC per base sequence quality ITS Compost Robin

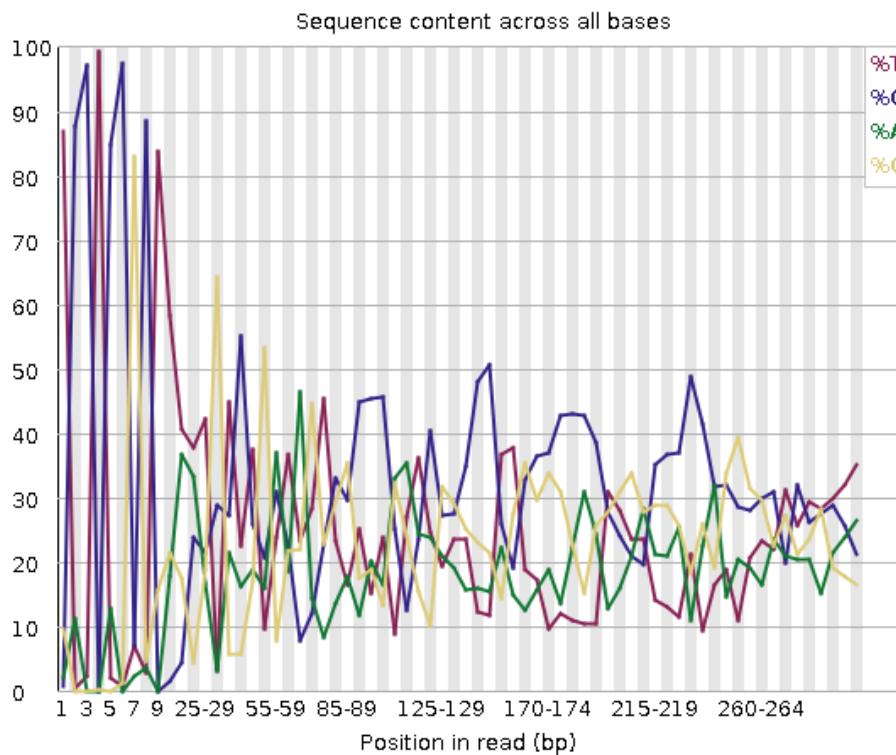


Figure 46 : FASTQC quality per base sequence content R1 ITS Compost Robin

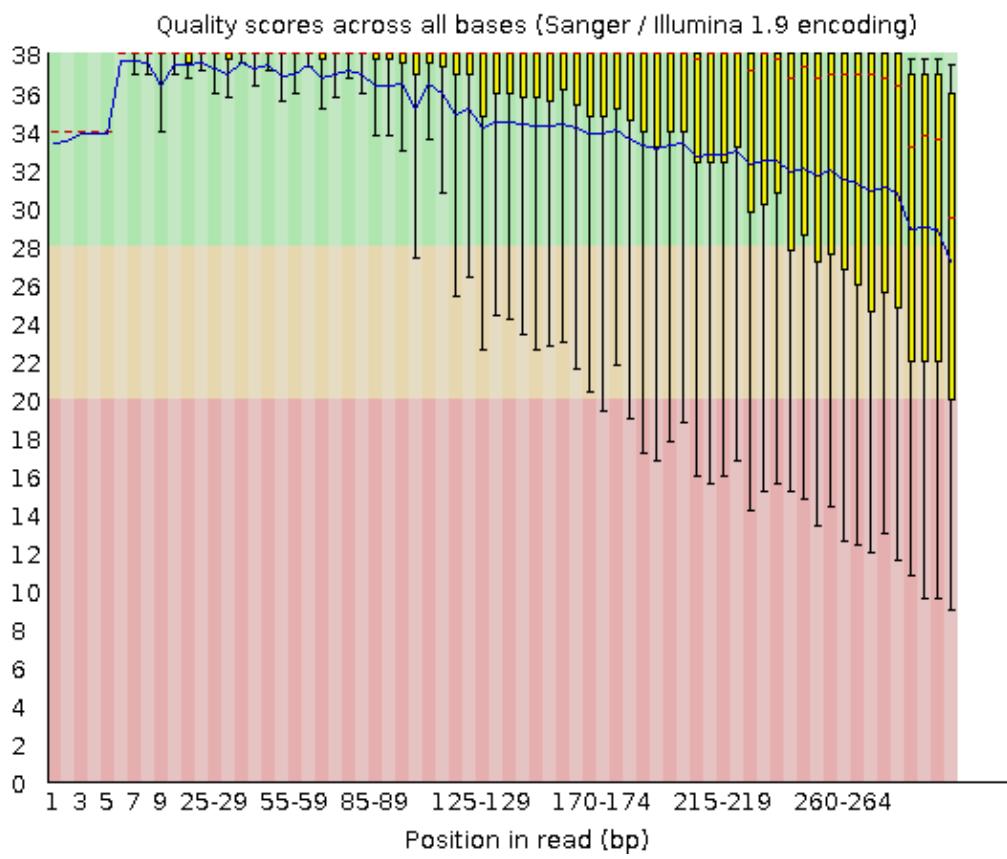


Figure 47 FASTQC R1 per base sequence quality 16S Compost Robin

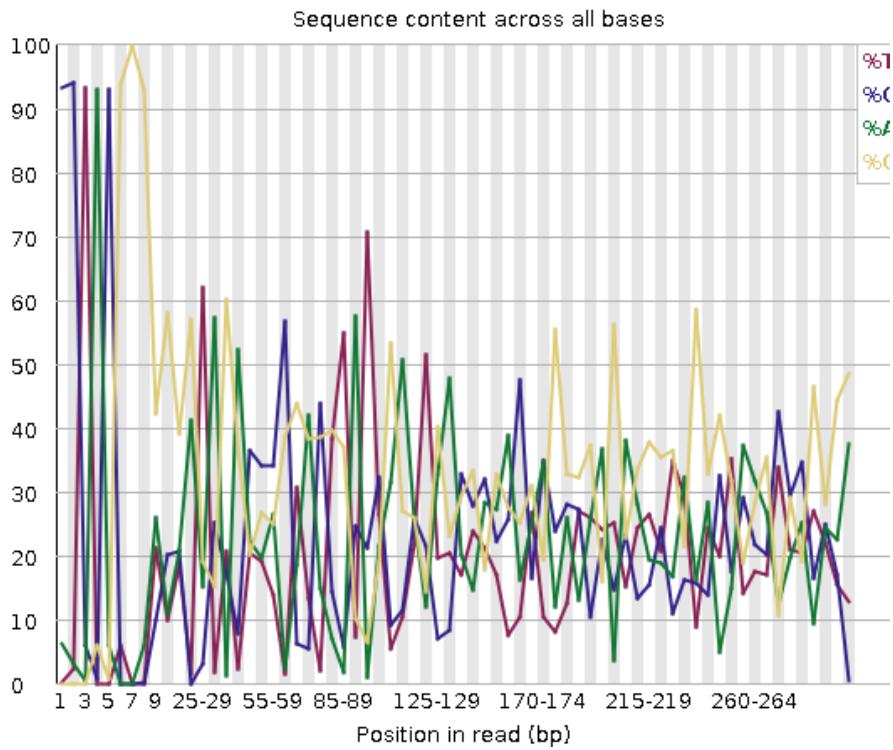


Figure 48 FASTQC R1 per base sequence content 16S Compost Robin

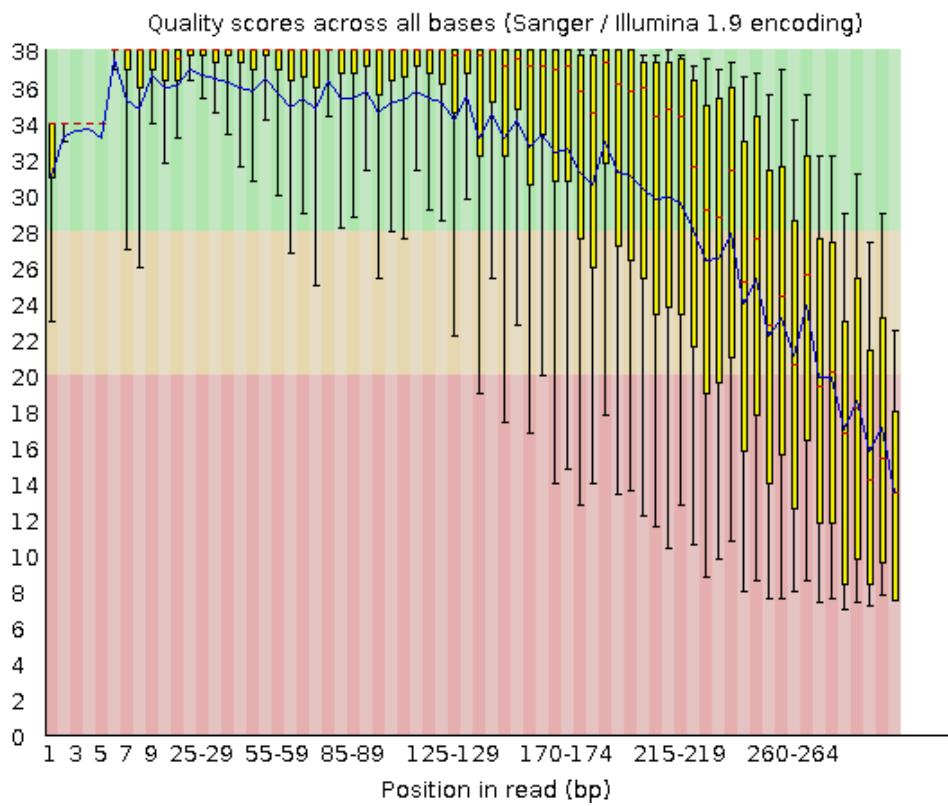


Figure 49 FASTQC R2 per base sequence quality 16S Compost Robin

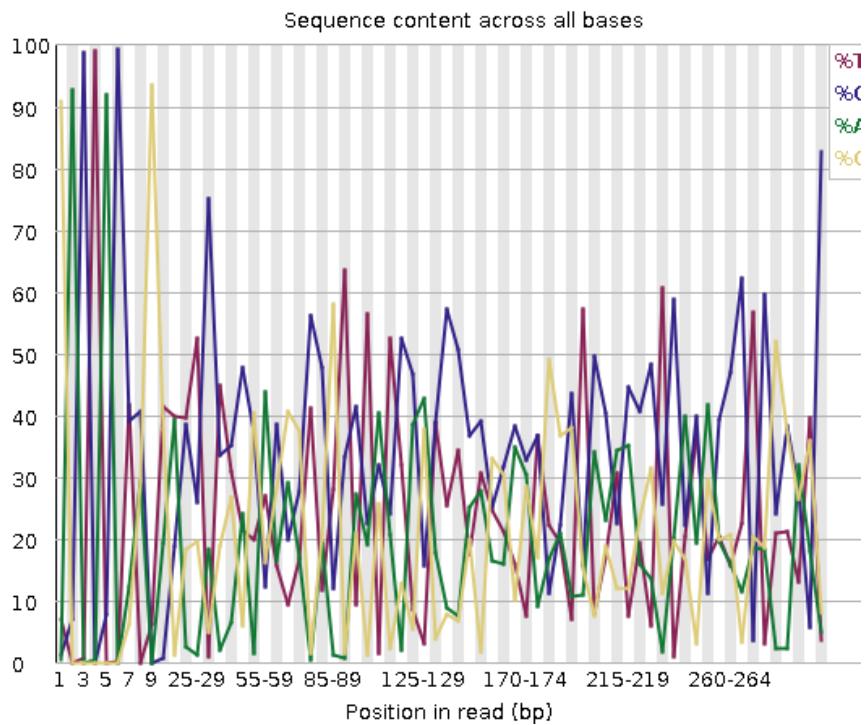


Figure 50 FASTQC R2 per base sequence content 16S Compost Robin

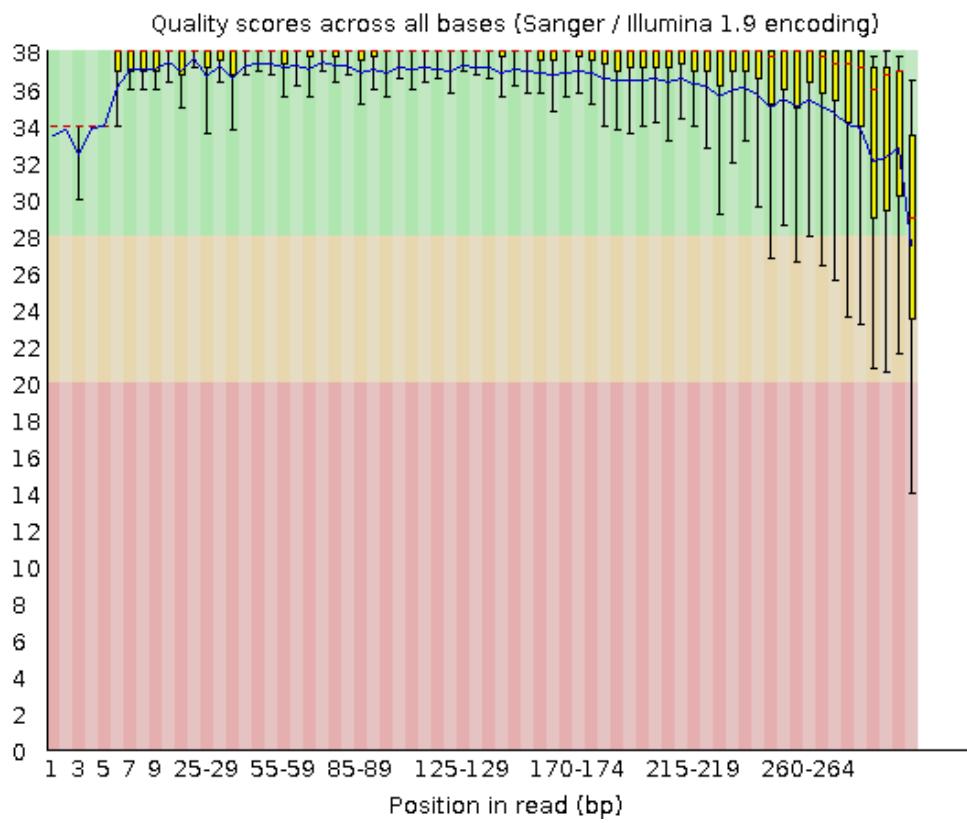


Figure 51 FASTQC R2 per base sequence quality 18S Compost Robin

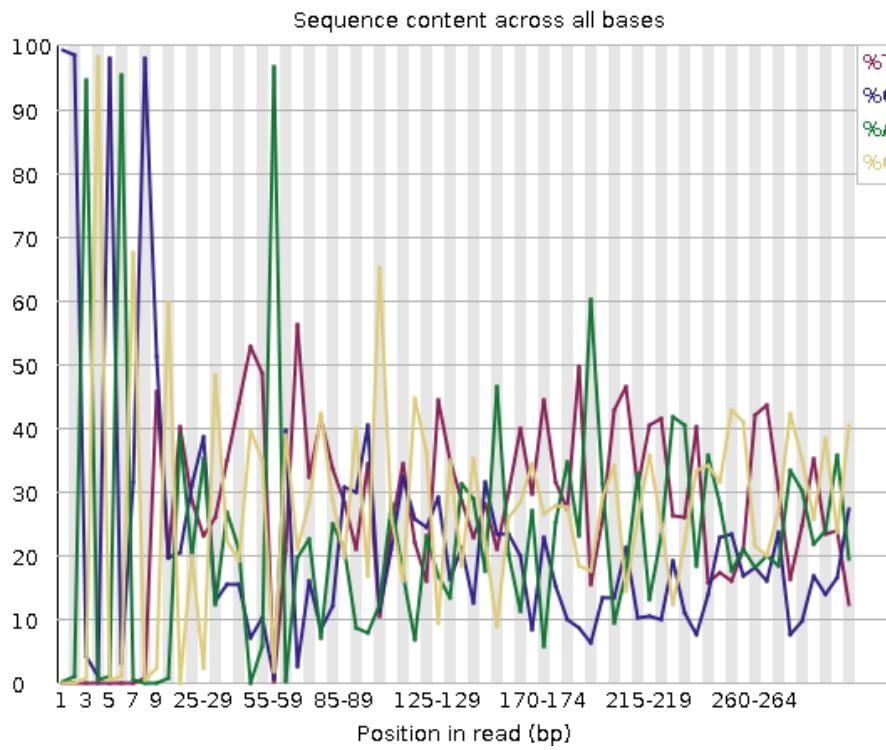


Figure 52 FASTQC R2 per base sequence content 18S Compost Robin

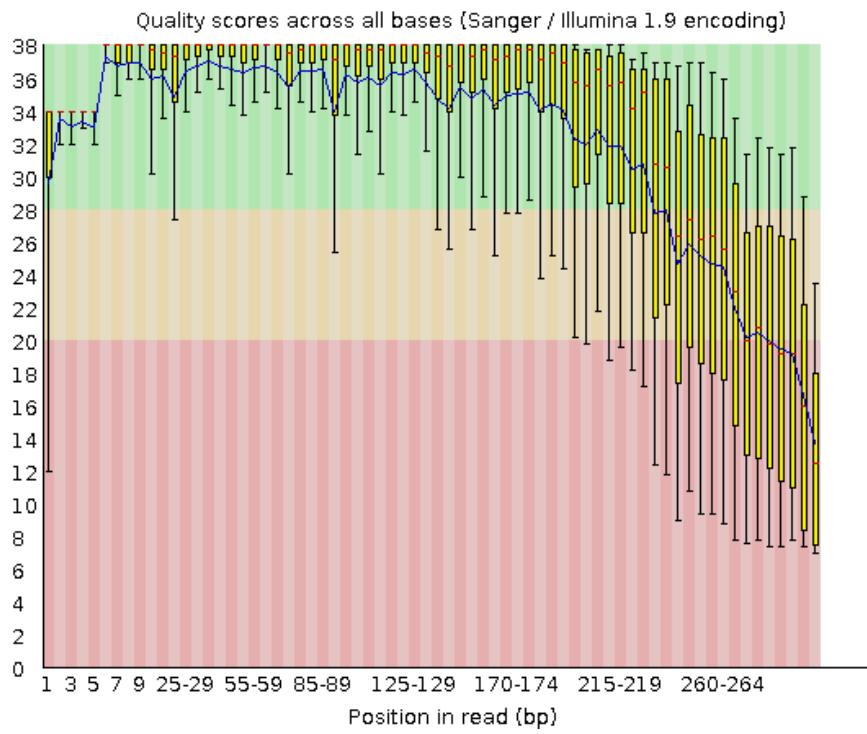


Figure 53 FASTQC R1 per base sequence quality 18S Compost Robin

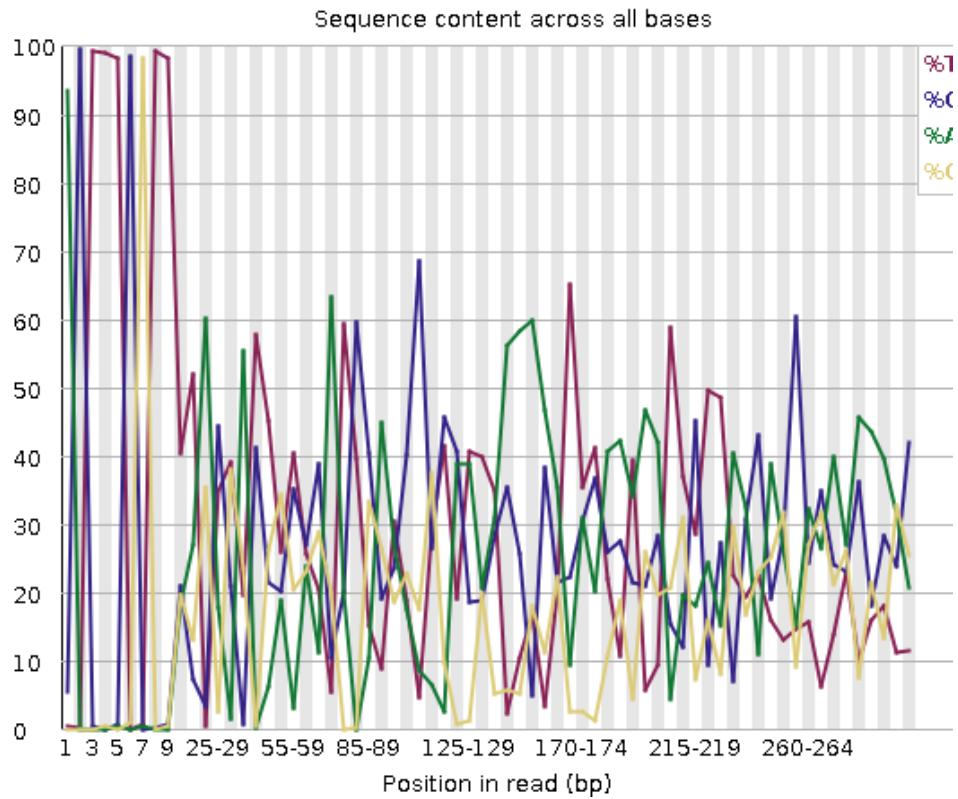


Figure 54 FASTQC R1 per base sequence content 18S Compost Robin

7.1.2 FASTQC Guillaume Eau

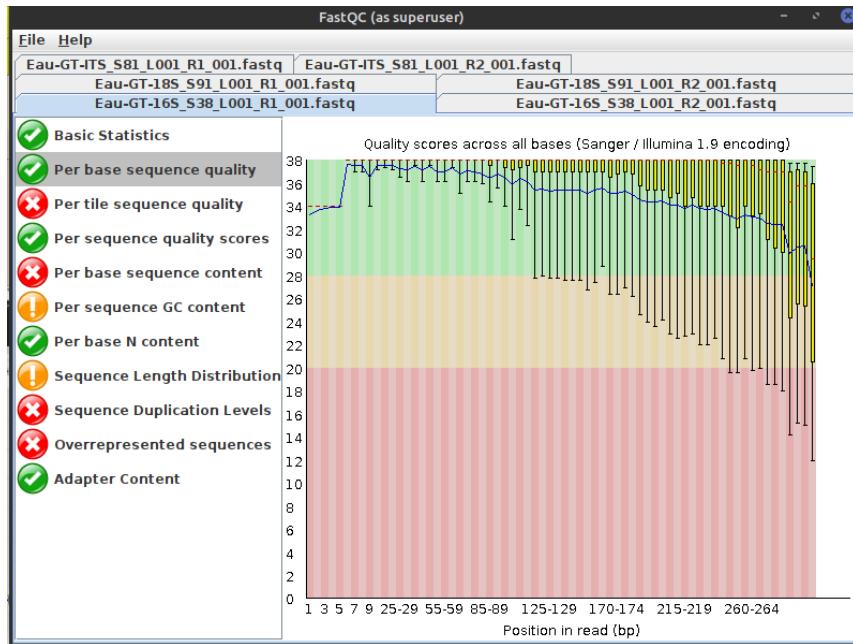


Figure 55 FASTQC quality per base sequence quality R1 16S Eau

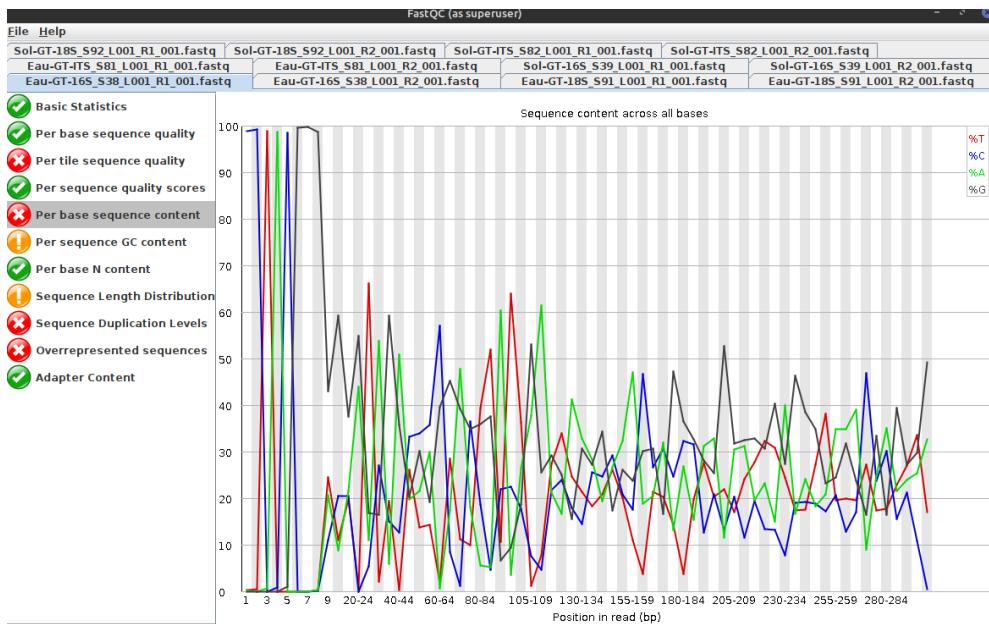


Figure 56 FASTQC quality per base sequence content R1 16S Eau

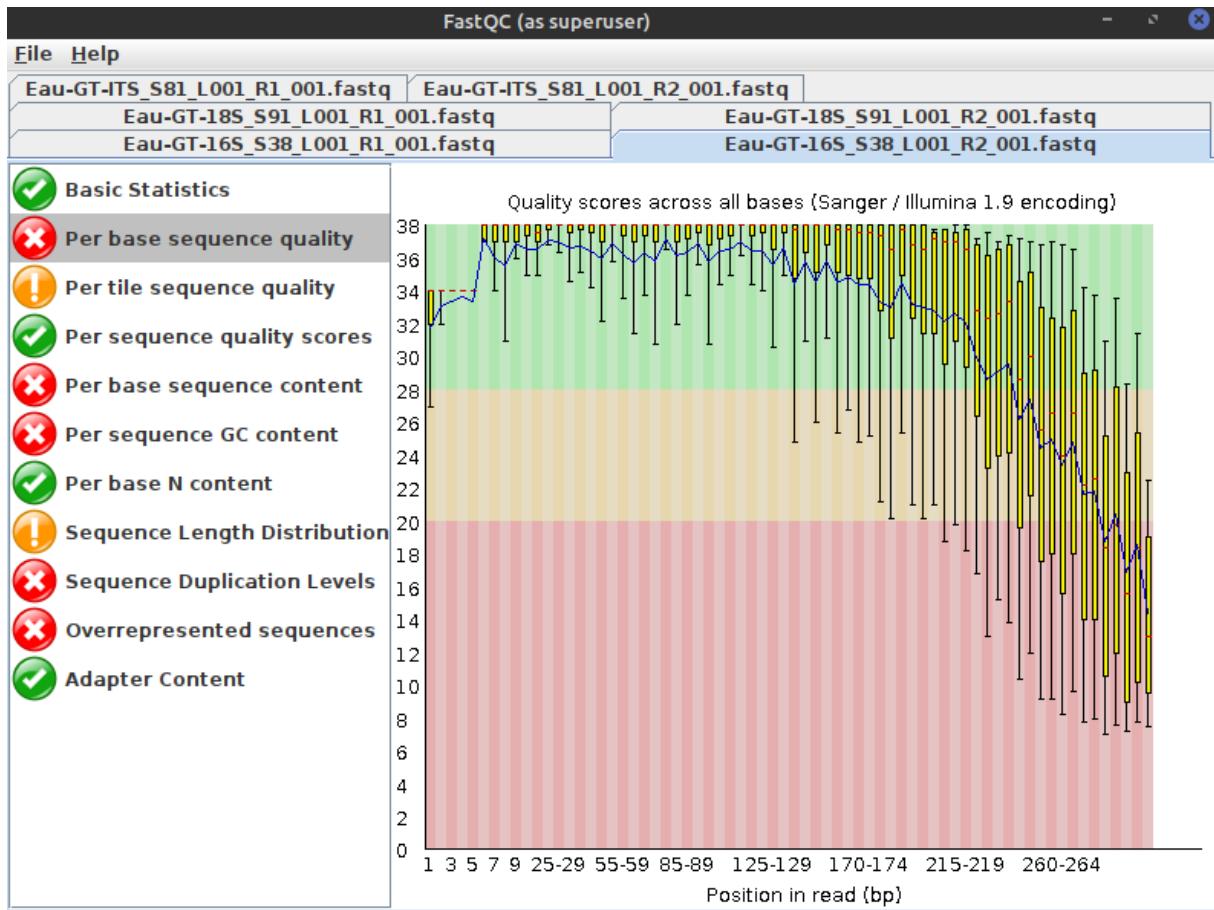


Figure 57 FASTQC quality per base sequence quality R2 16S Eau

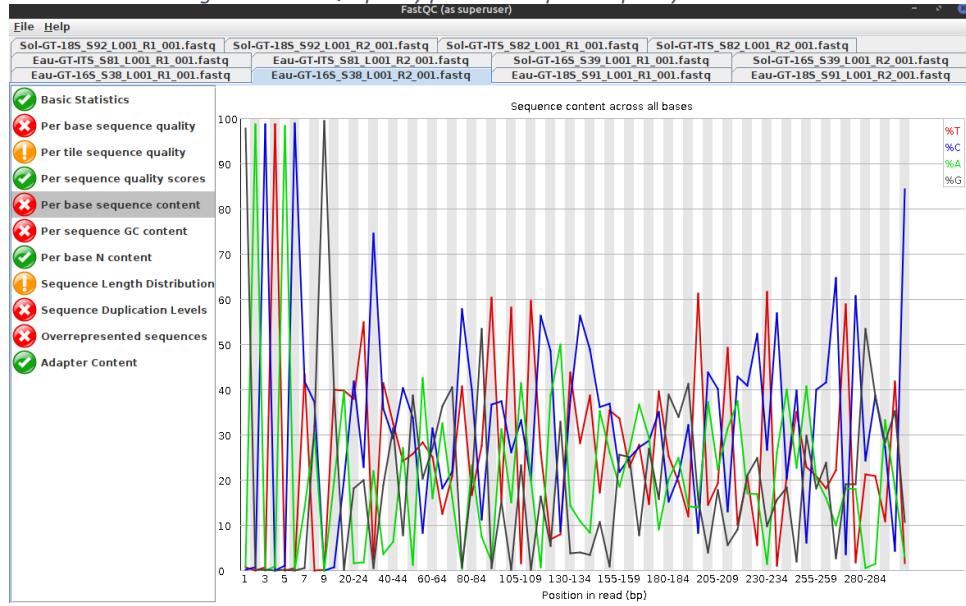


Figure 58 FASTQC quality per base sequence content R2 16S Eau

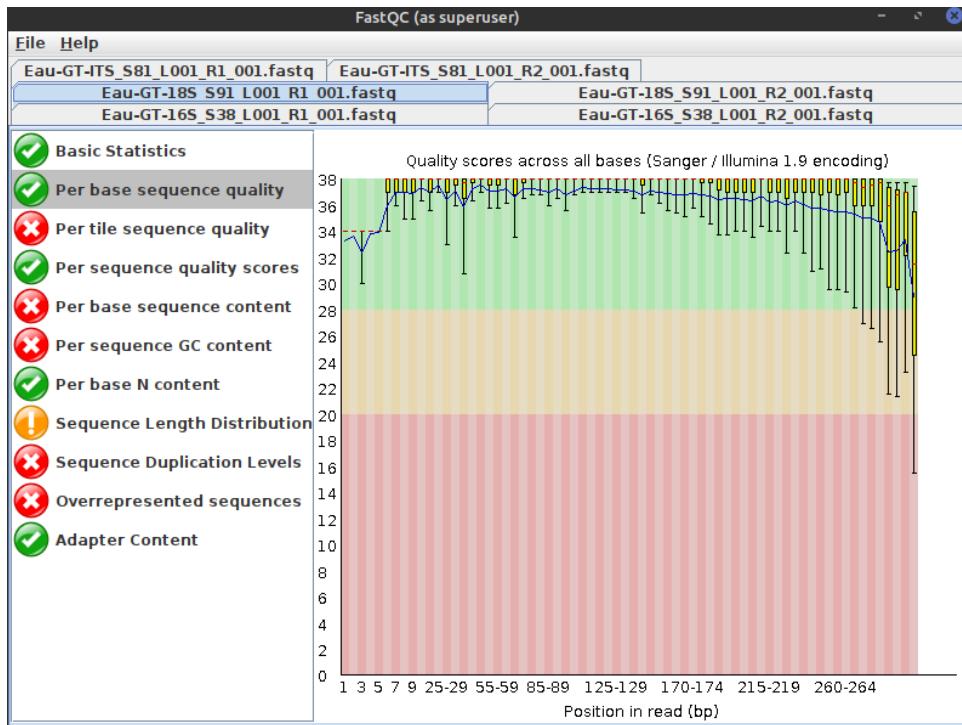


Figure 59 FASTQC quality per base sequence quality R1 18S Eau

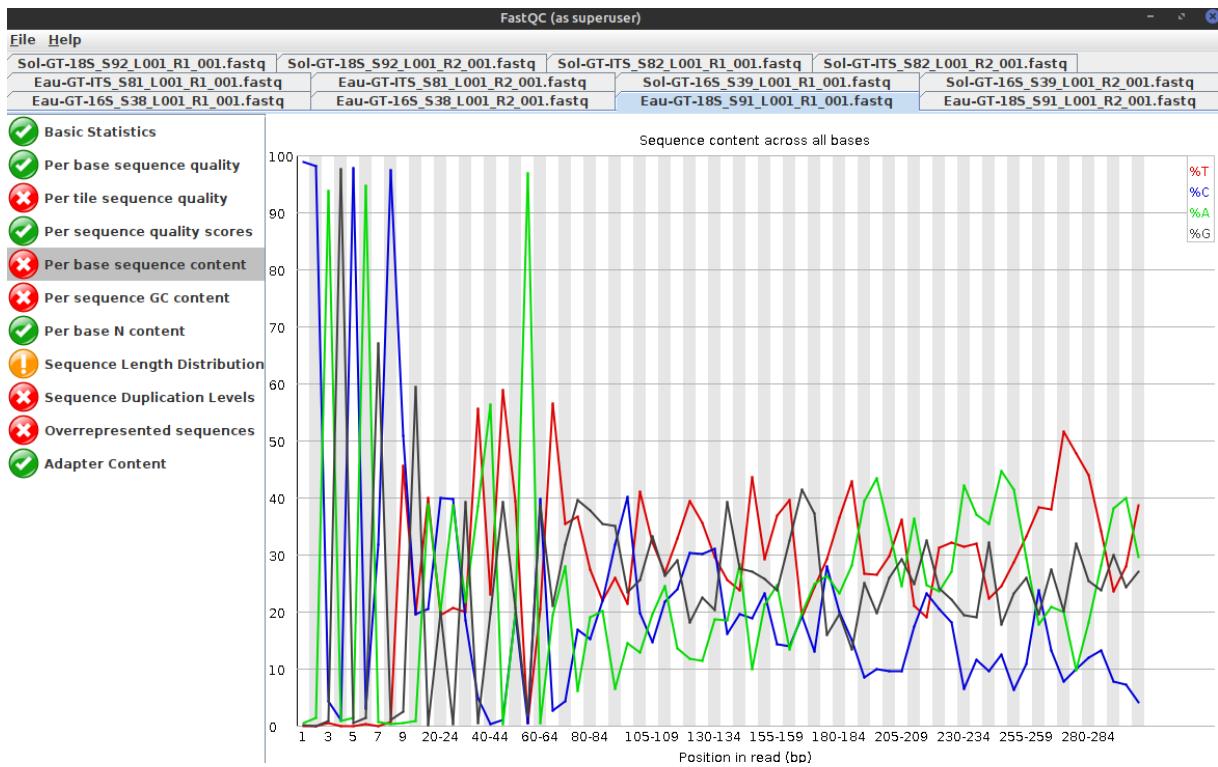


Figure 60 FASTQC quality per base sequence content R1 18S Eau

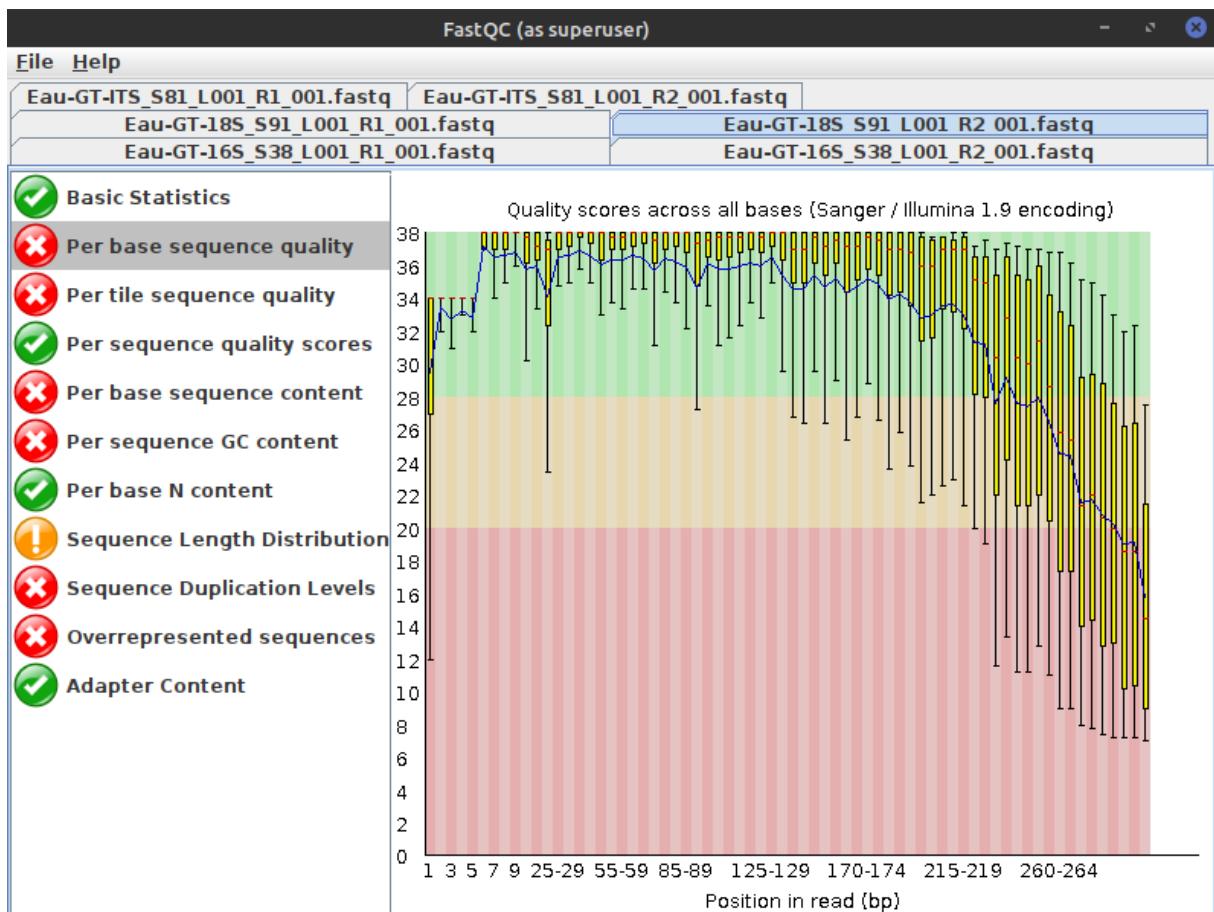


Figure 61 FASTQC quality per base sequence quality R2 18S Eau

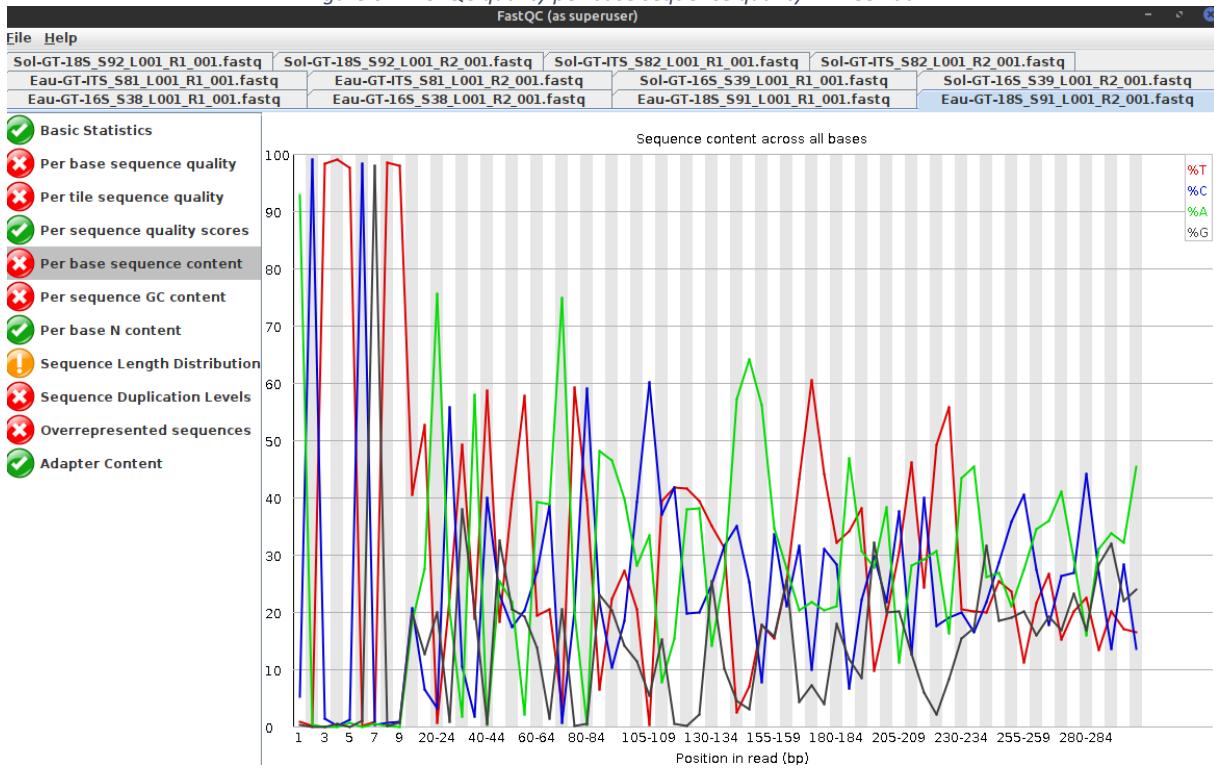


Figure 62 FASTQC quality per base sequence content R2 18S Eau

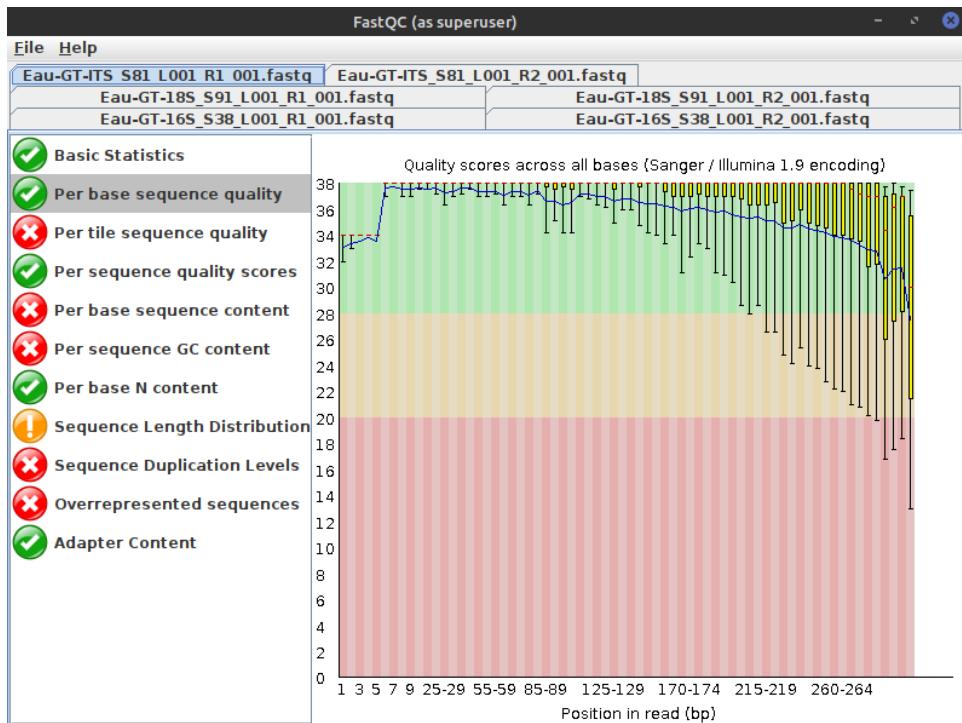


Figure 63 FASTQC quality per base sequence quality R1 ITS Eau

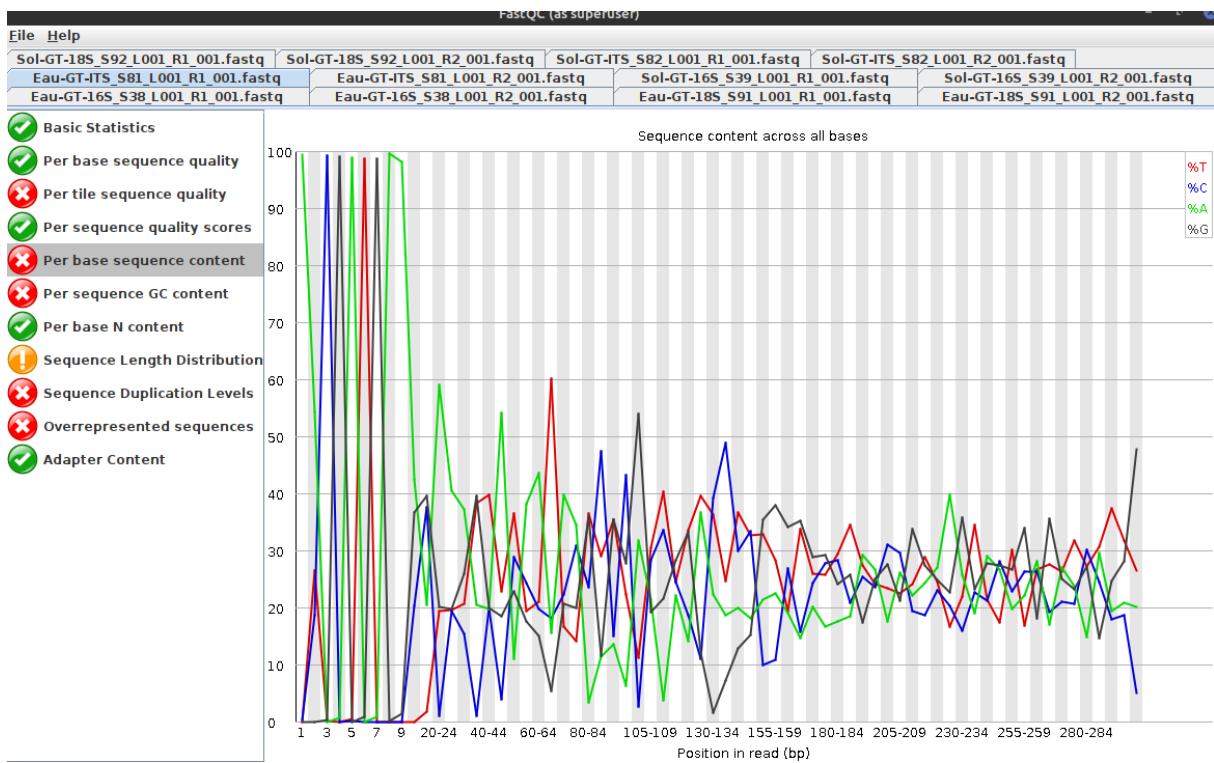


Figure 64 FASTQC quality per base sequence content R1 ITS Eau

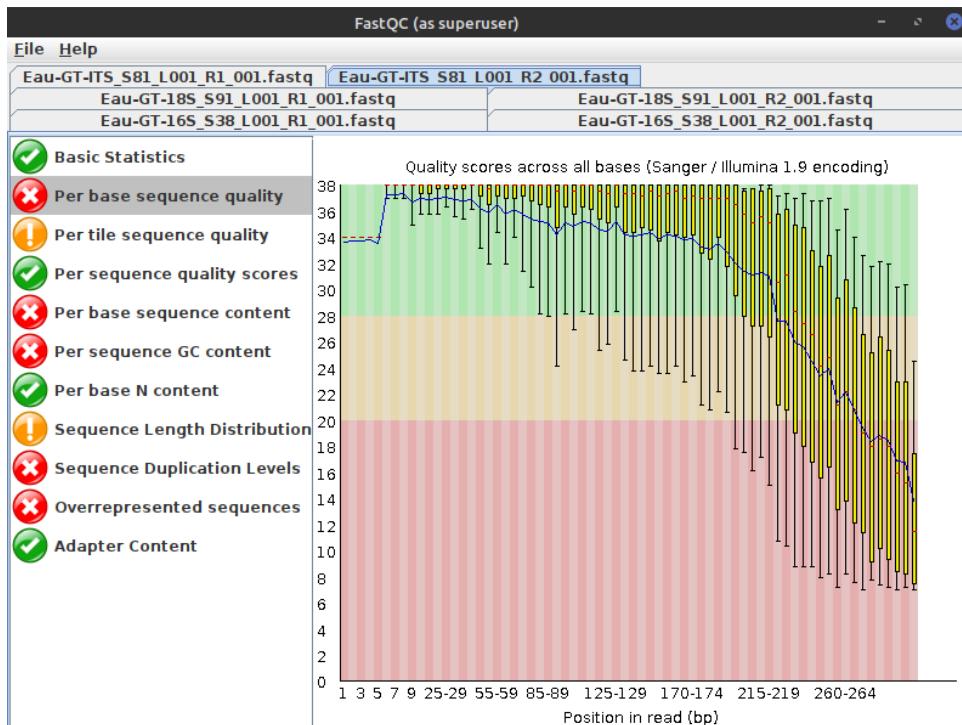


Figure 65 FASTQC quality per base sequence quality R ITS Eau

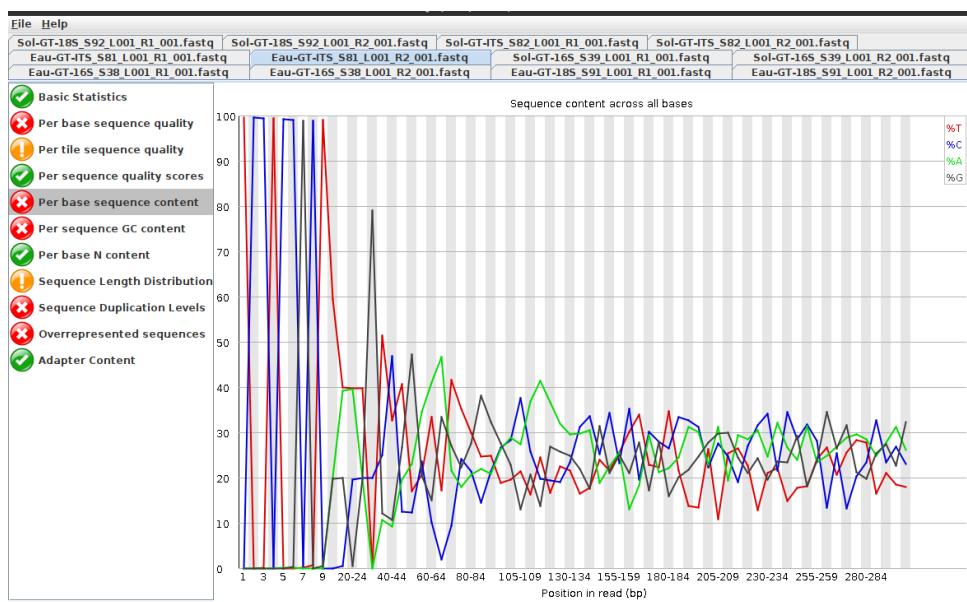


Figure 66 FASTQC quality per base sequence content R2 ITS Eau

7.1.3 FASTQC Guillaume Terre

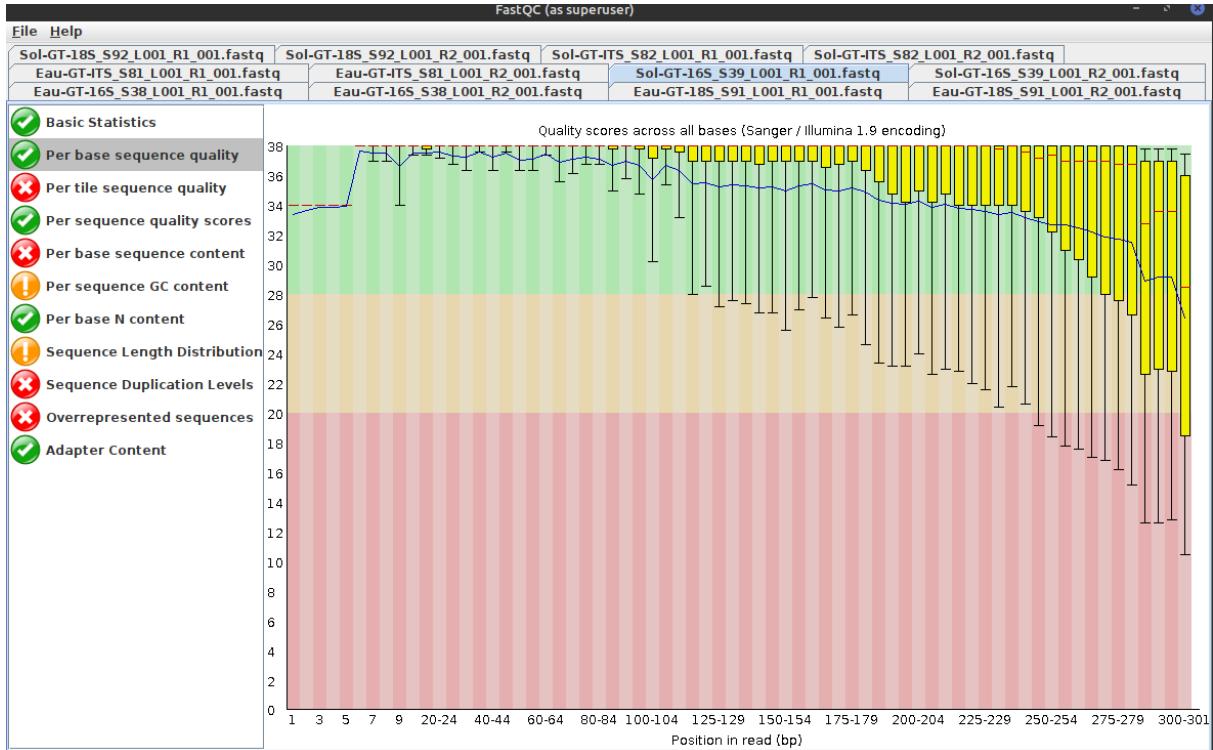


Figure 67 16S per base sequence quality R1 Sol

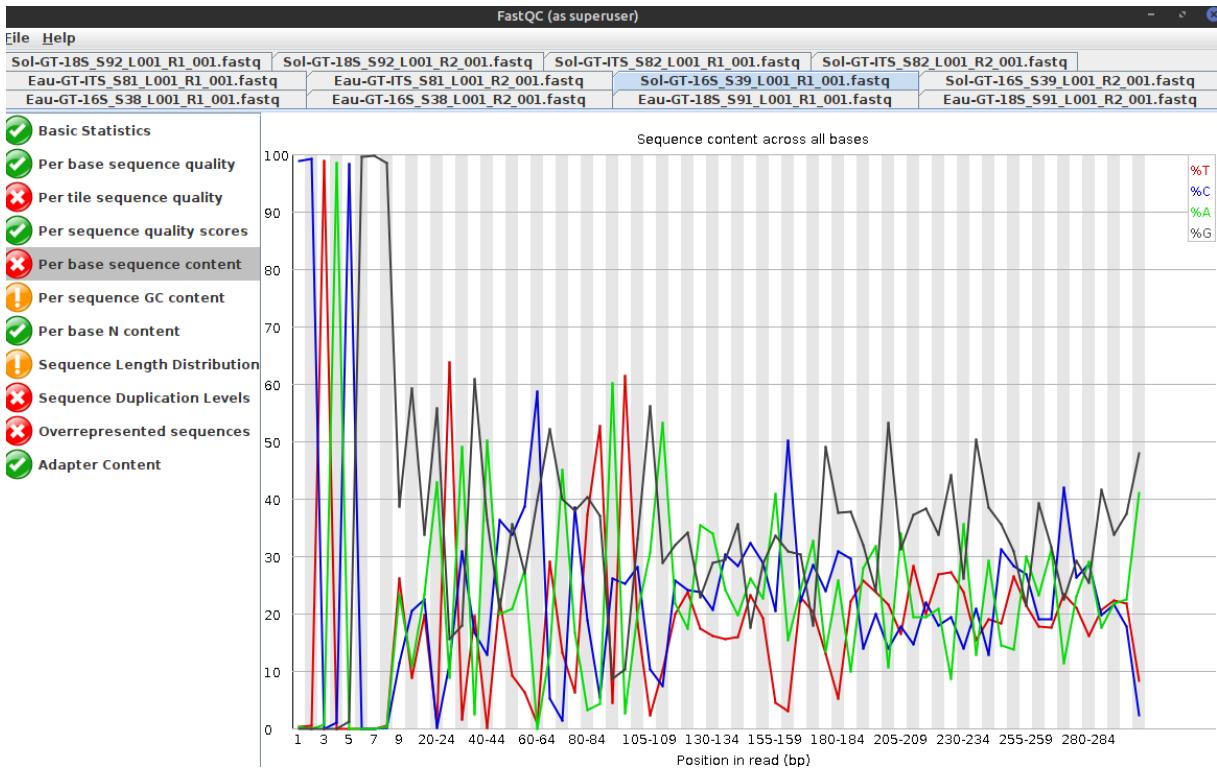


Figure 68 16S per base sequence content R1 Sol

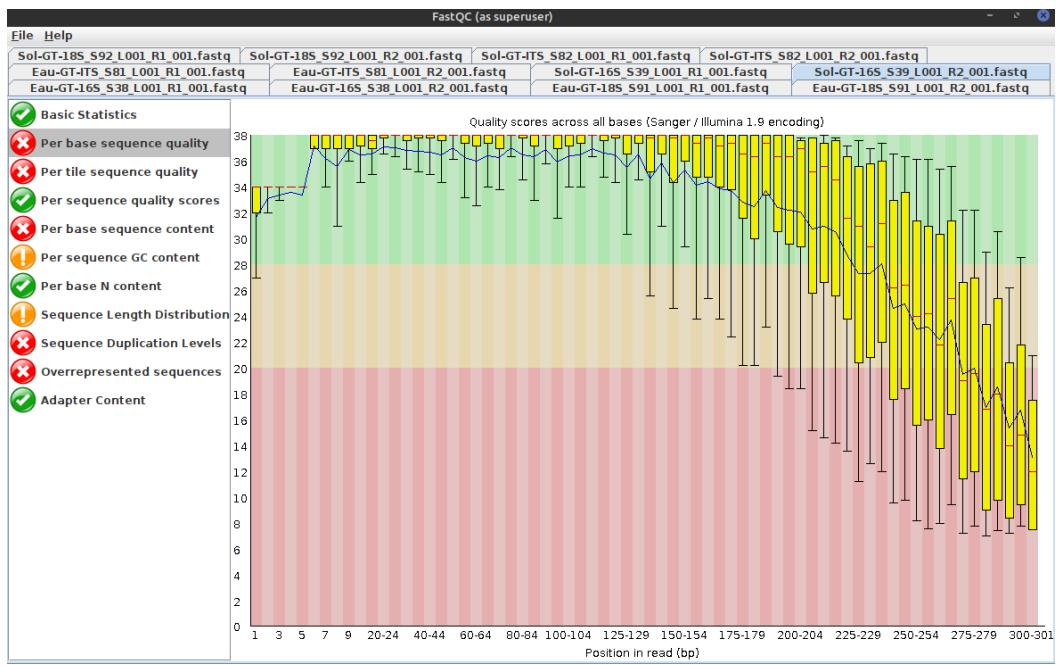


Figure 69 16S per base sequence quality R2 Sol

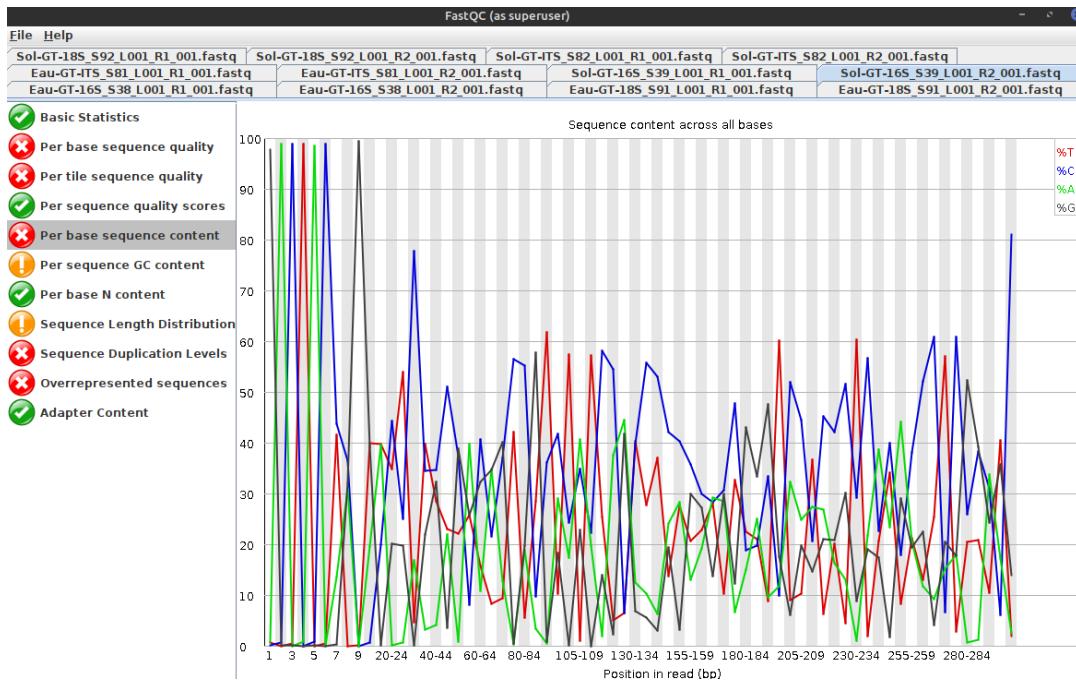


Figure 70 16S per base sequence content R2 Sol

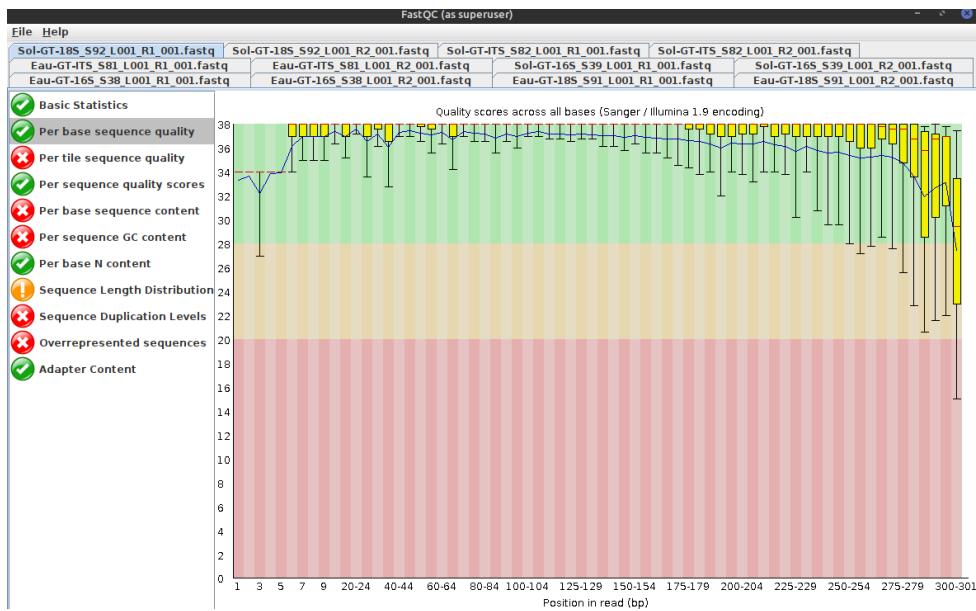


Figure 71 18S per base sequence quality R1 Sol

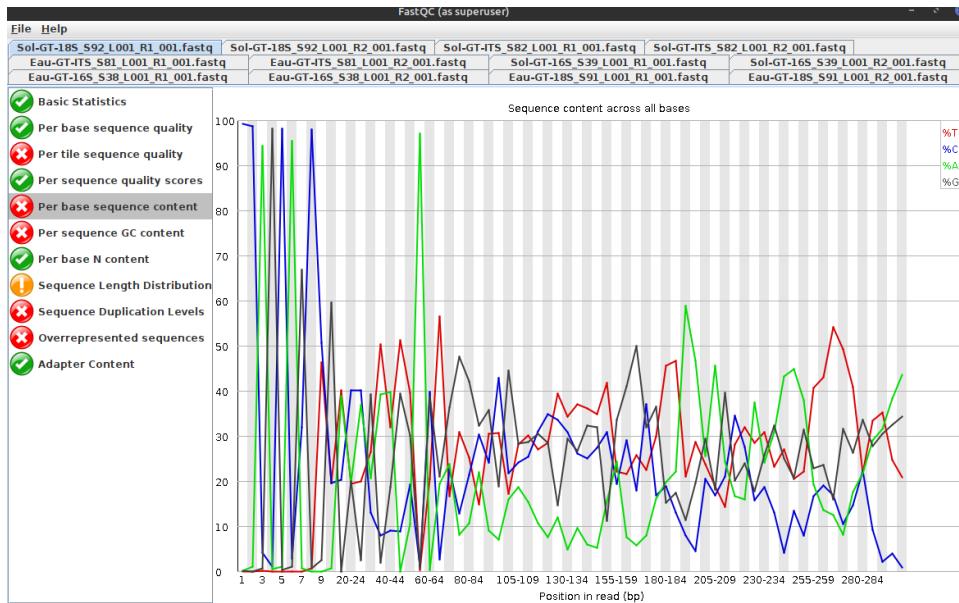


Figure 72 18S per base sequence content R1 Sol

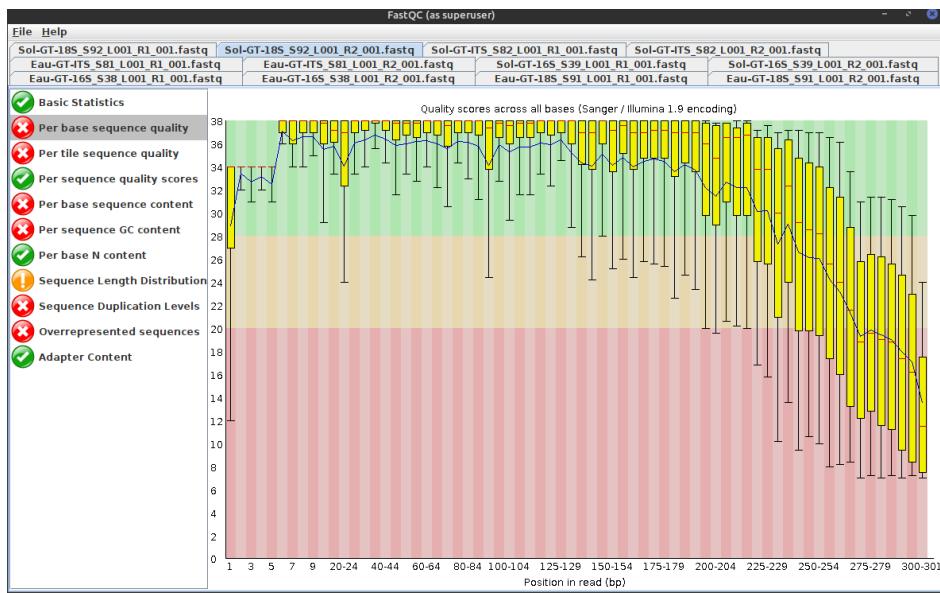


Figure 73 18S per base sequence quality R2 Sol

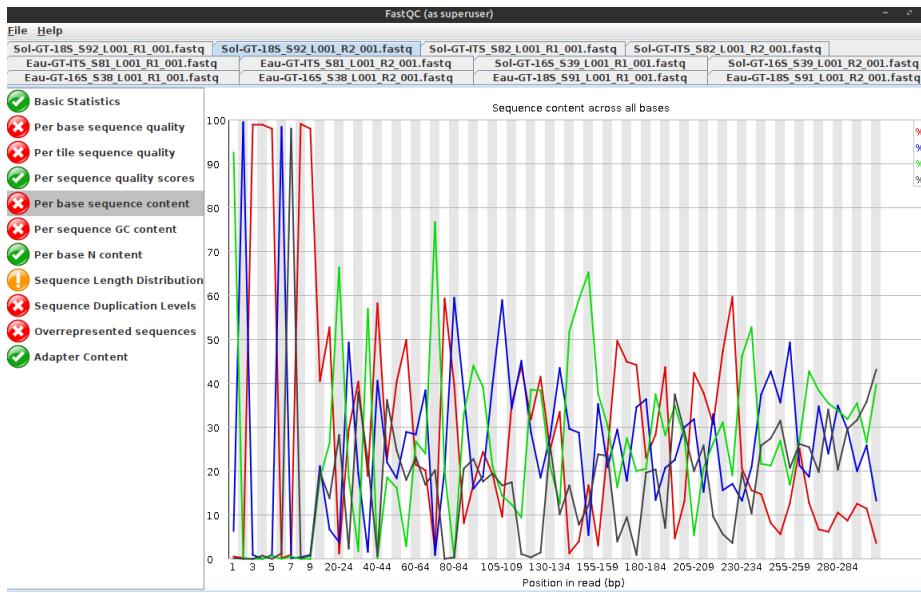


Figure 74 18S per base sequence content R2 Sol

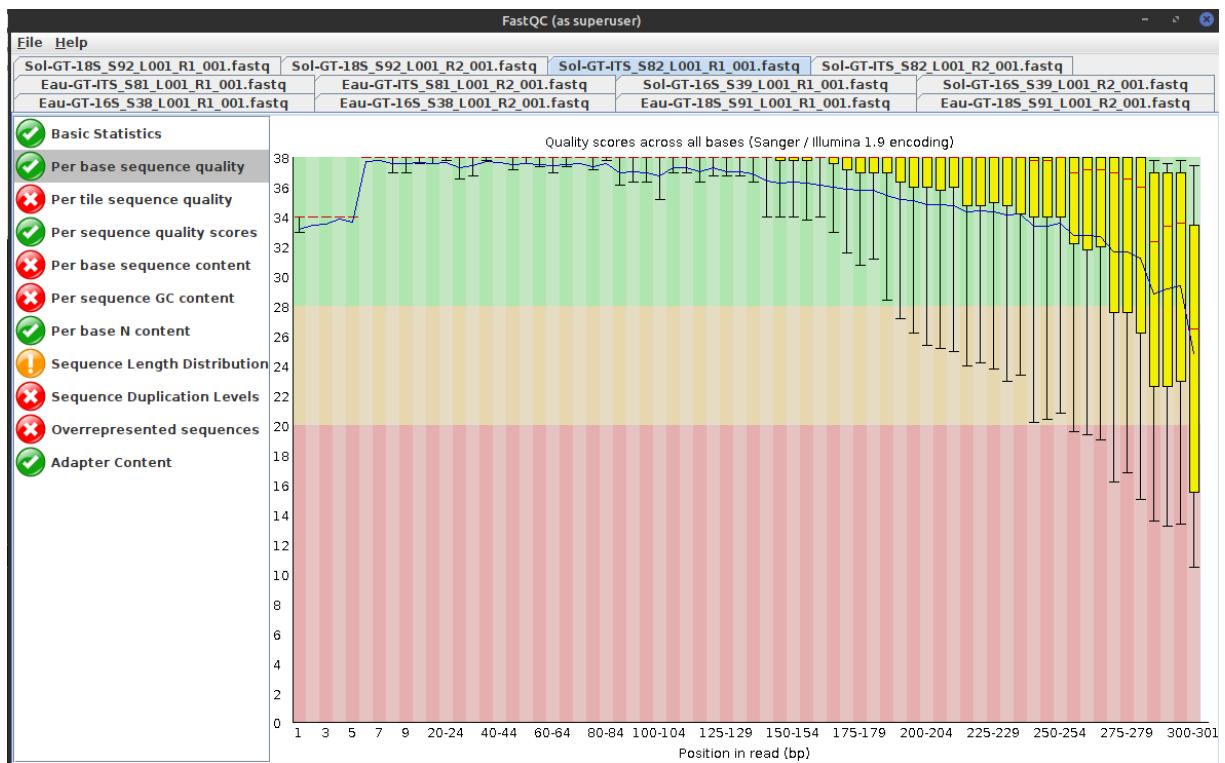


Figure 75 ITS per base sequence quality R1 Sol

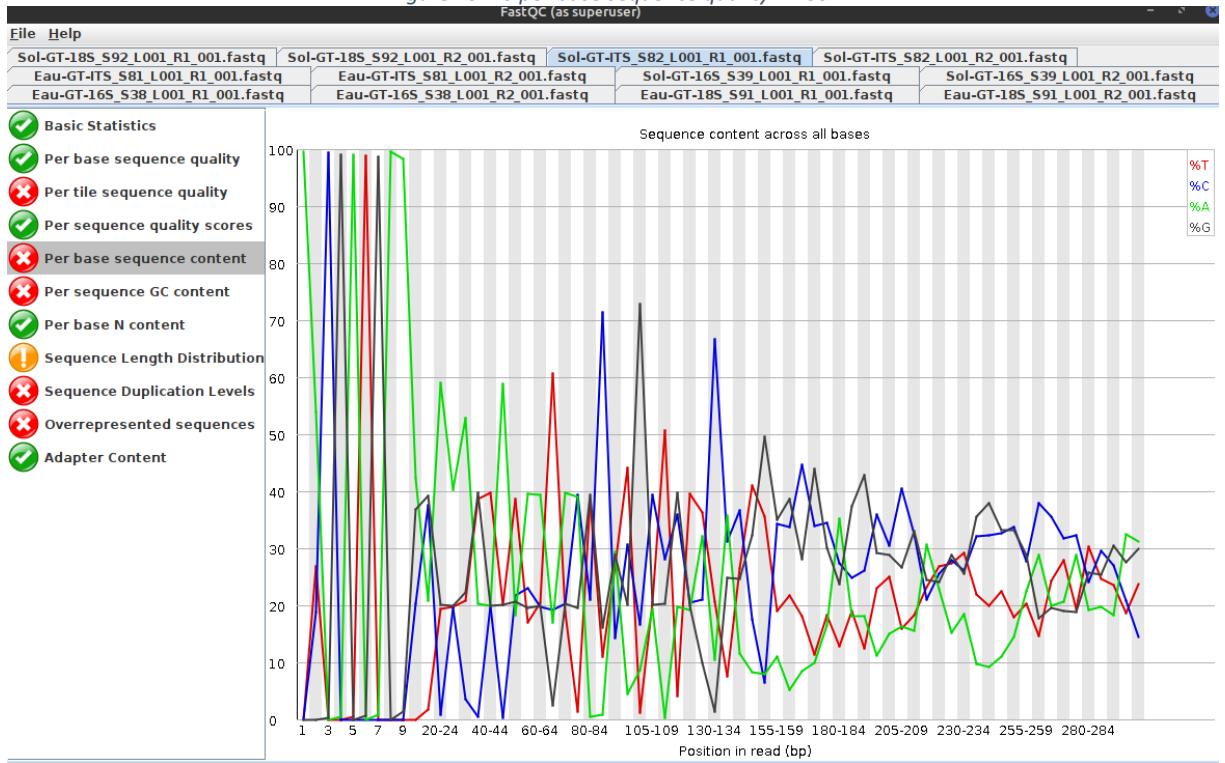


Figure 76 ITS per base sequence content R1 Sol

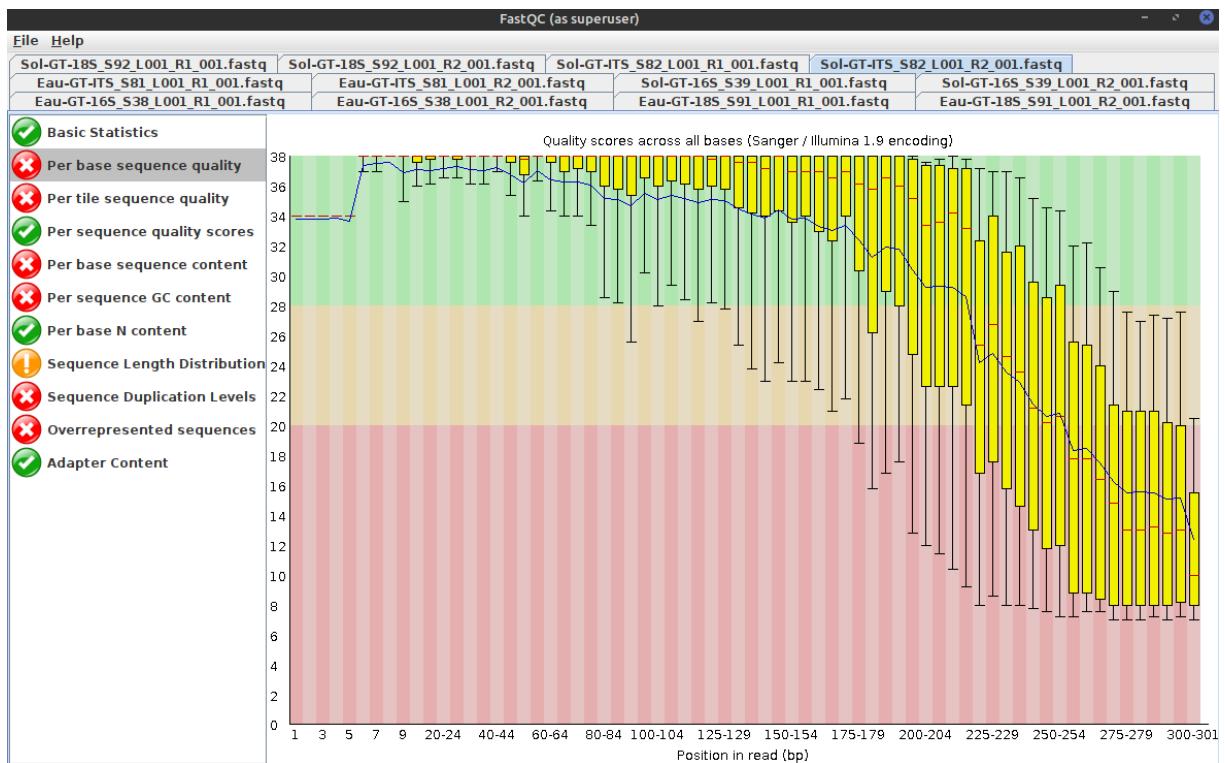


Figure 77 ITS per base sequence quality R2 Sol

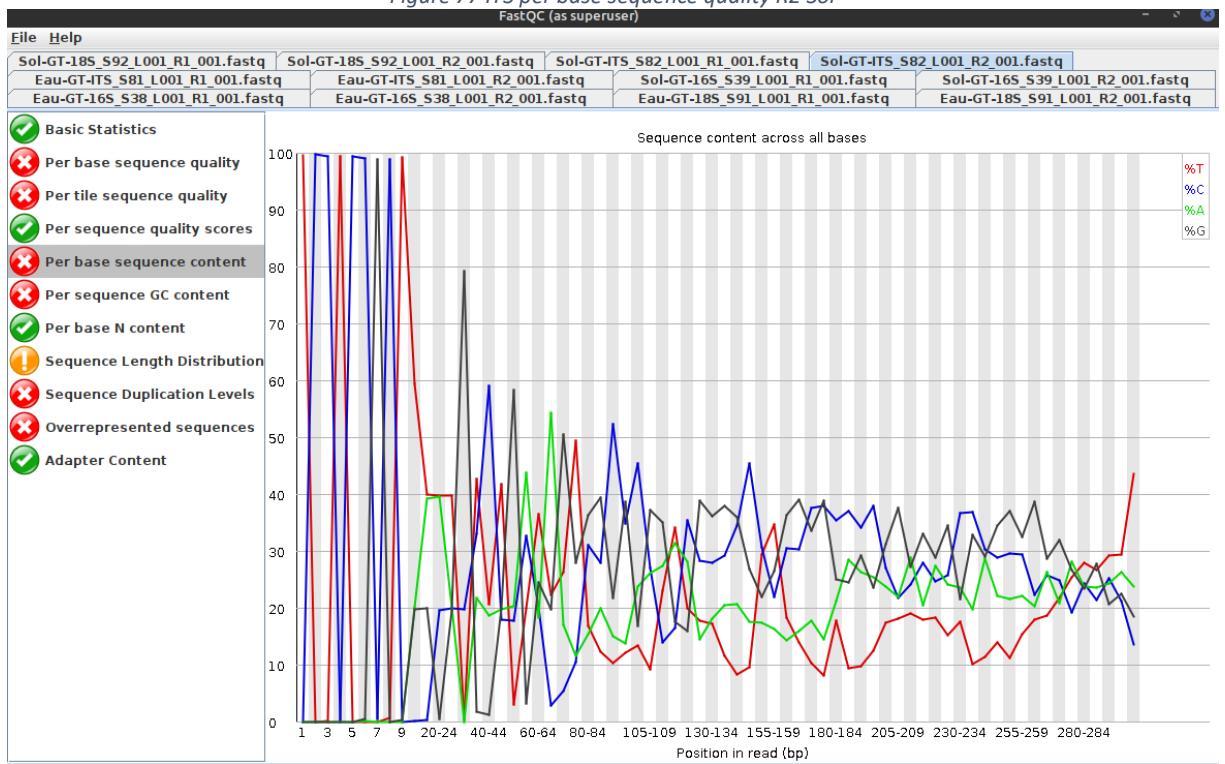


Figure 78 ITS per base sequence content R2 Sol

7.1.4 FASTQC Jus de patate

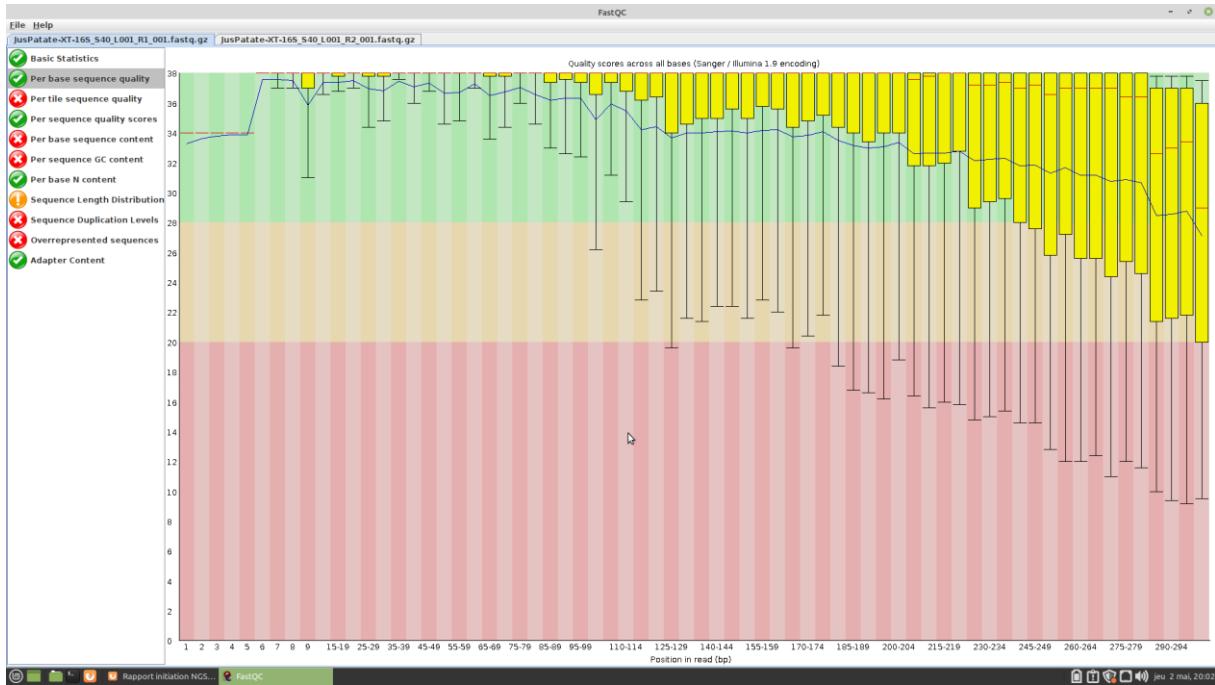


Figure 79 :16S per base sequence quality R1

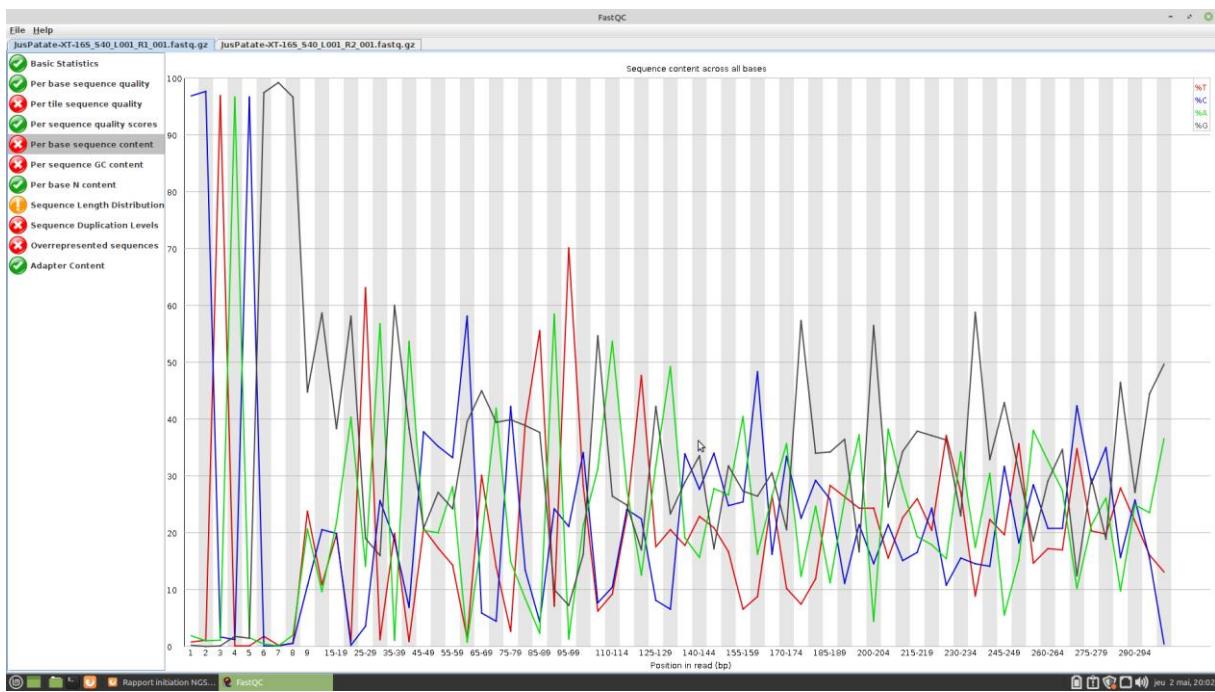


Figure 80 :16S per base sequence content R1

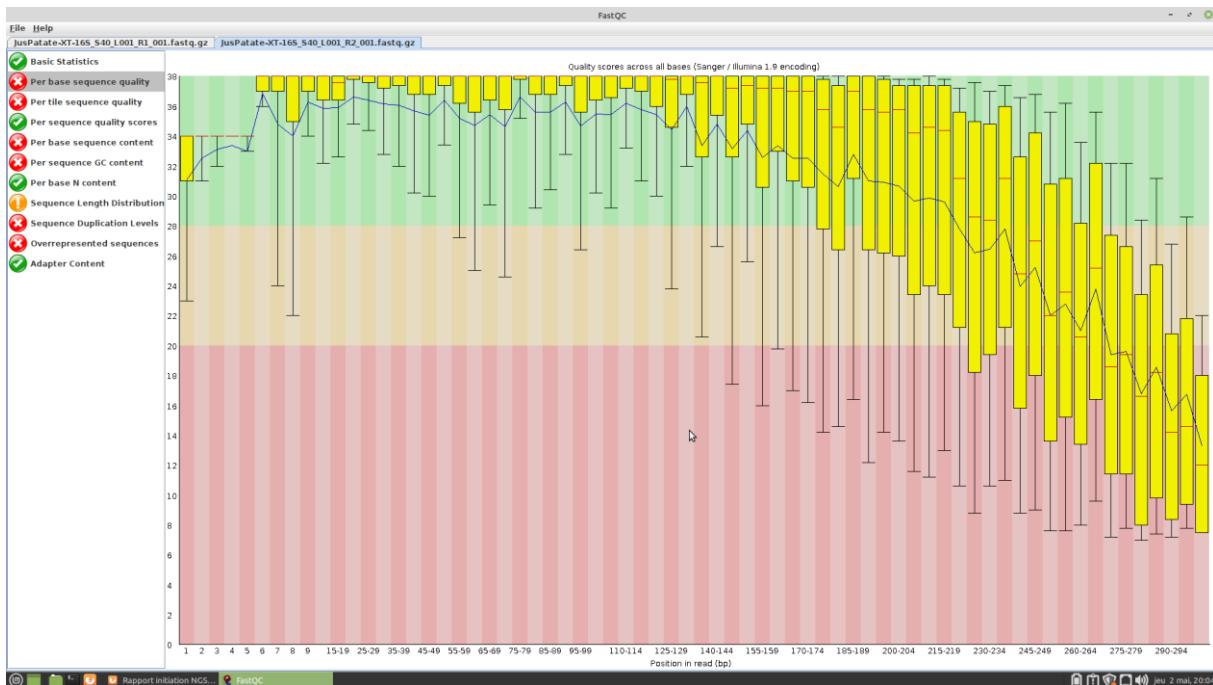


Figure 81 16S per base sequence quality R2 Jus de patate

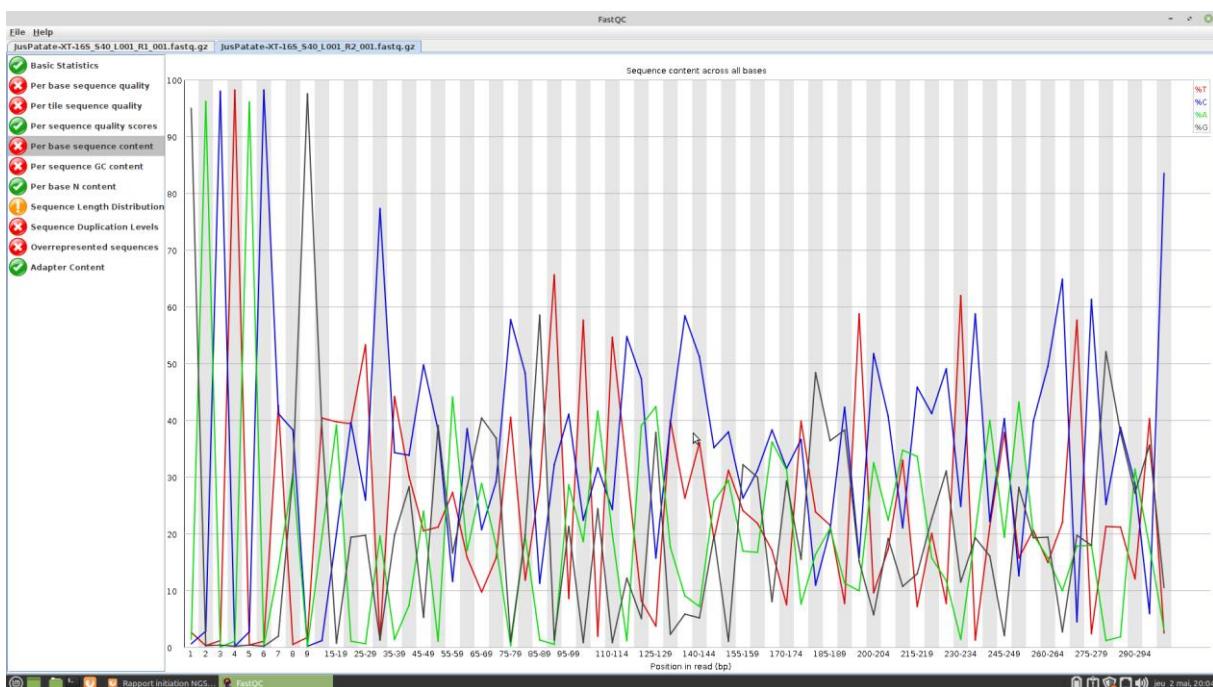


Figure 82 16S per base sequence content R2 Jus de patate

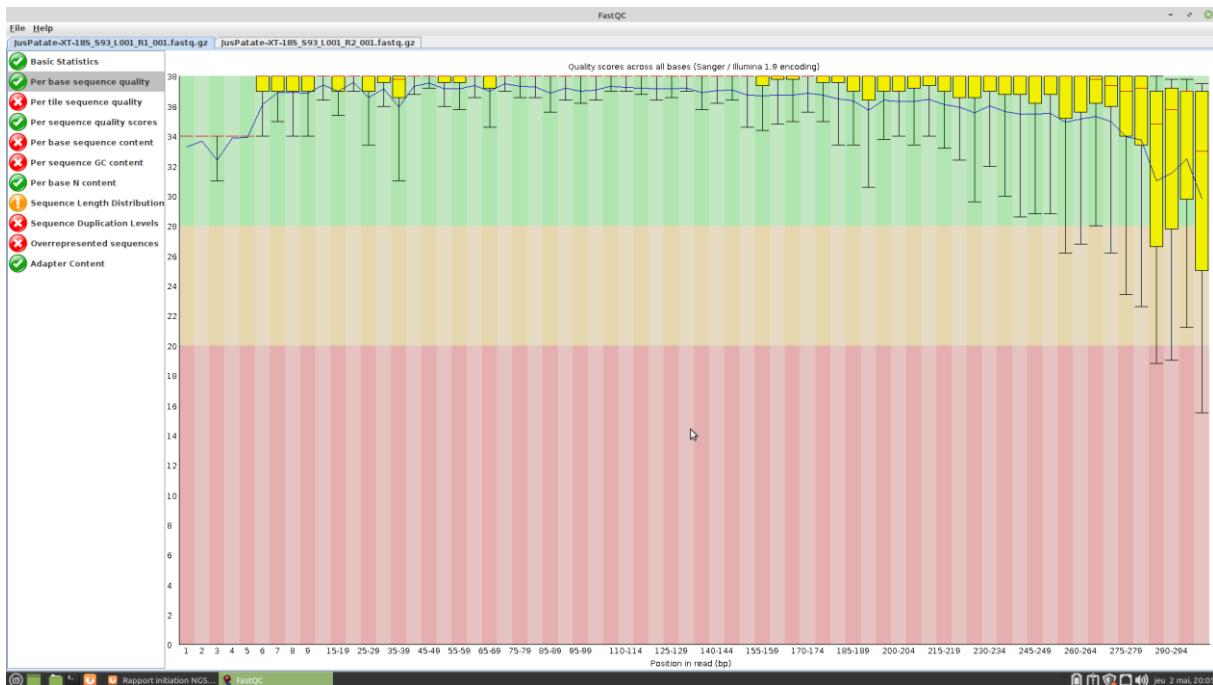


Figure 83 18S per base sequence quality R1 Jus de patate

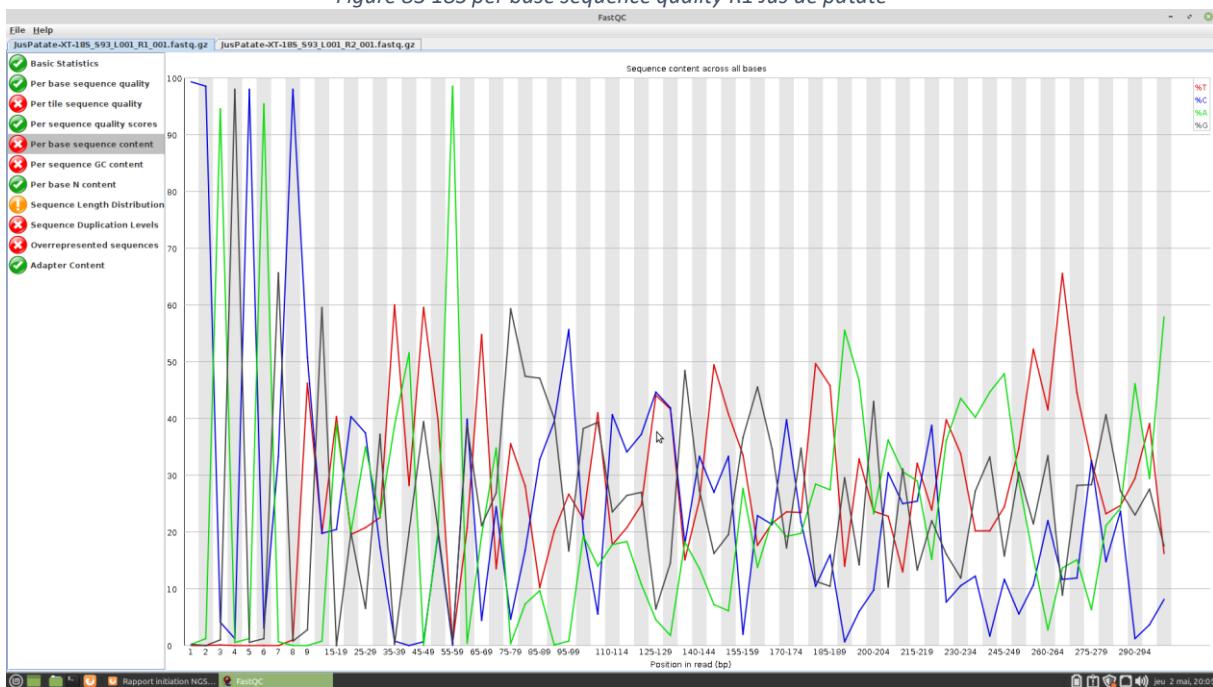


Figure 84 18S per base sequence content R1 Jus de patate

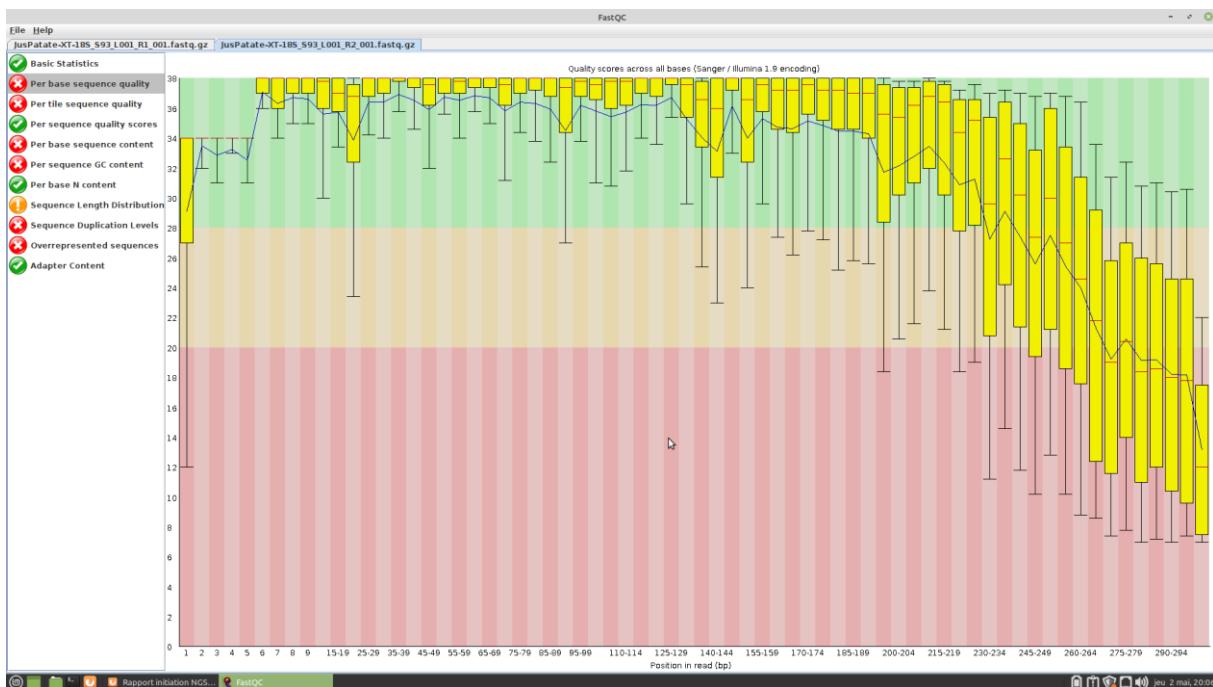


Figure 85 18S per base sequence quality R2 Jus de patate

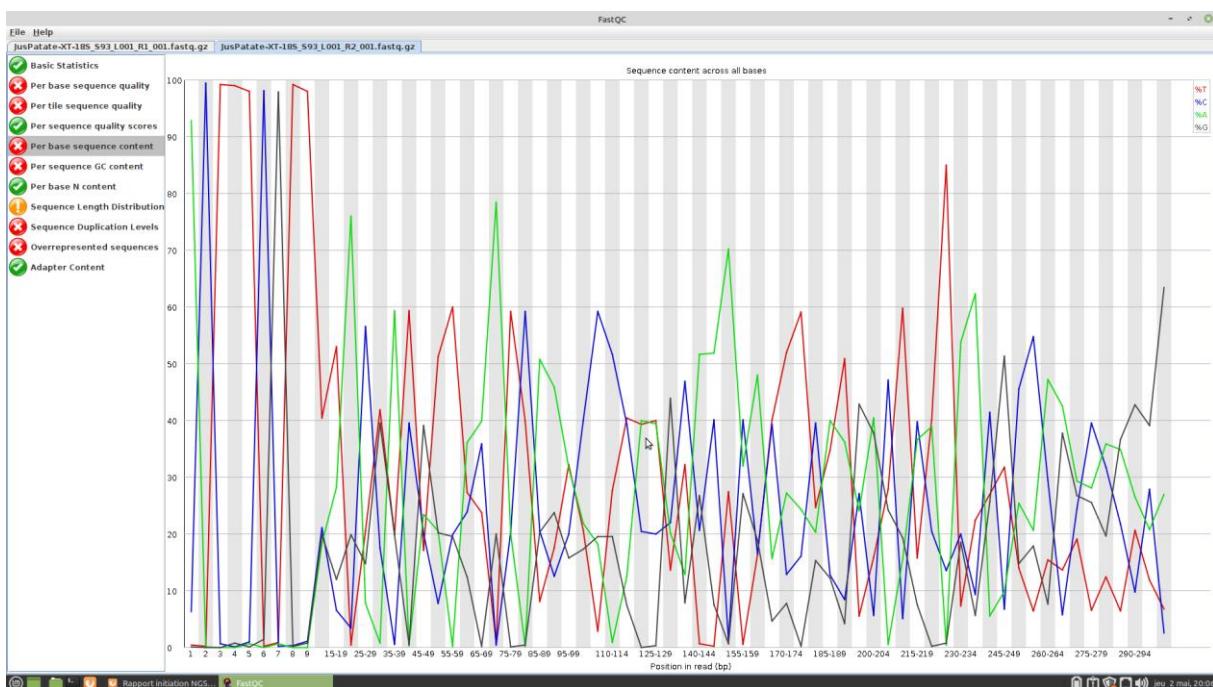


Figure 86 18S per base sequence content R2 Jus de patate

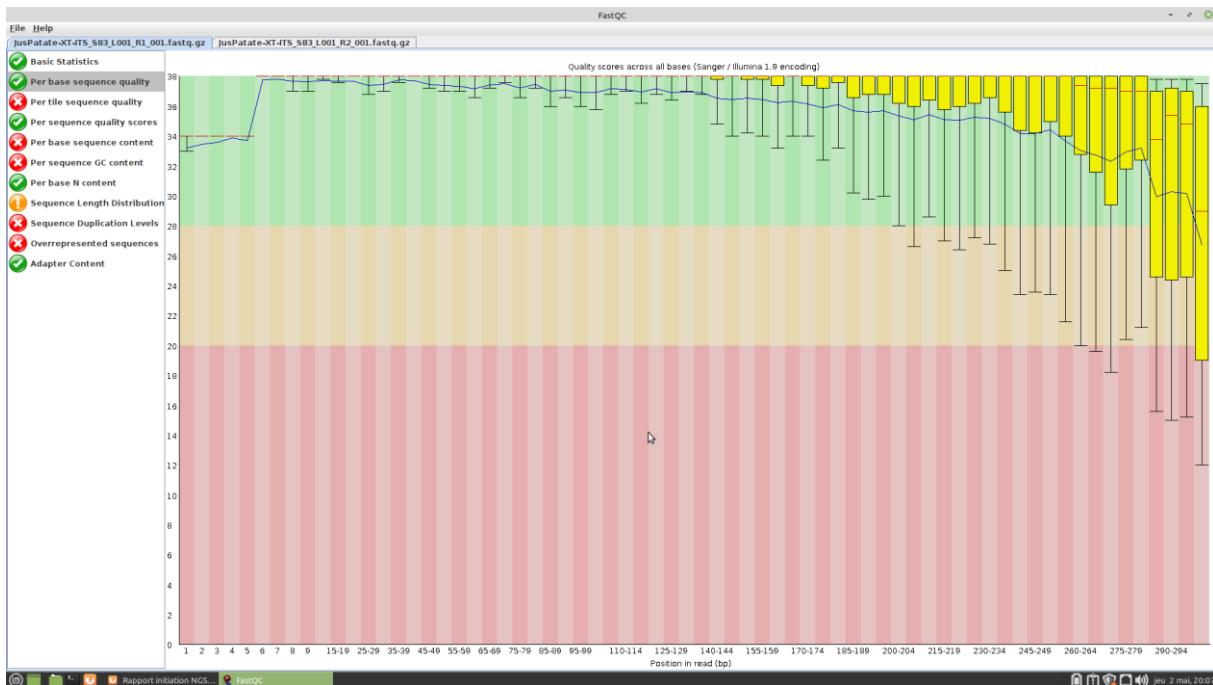


Figure 87 ITS per base sequence quality R1 Jus de patate

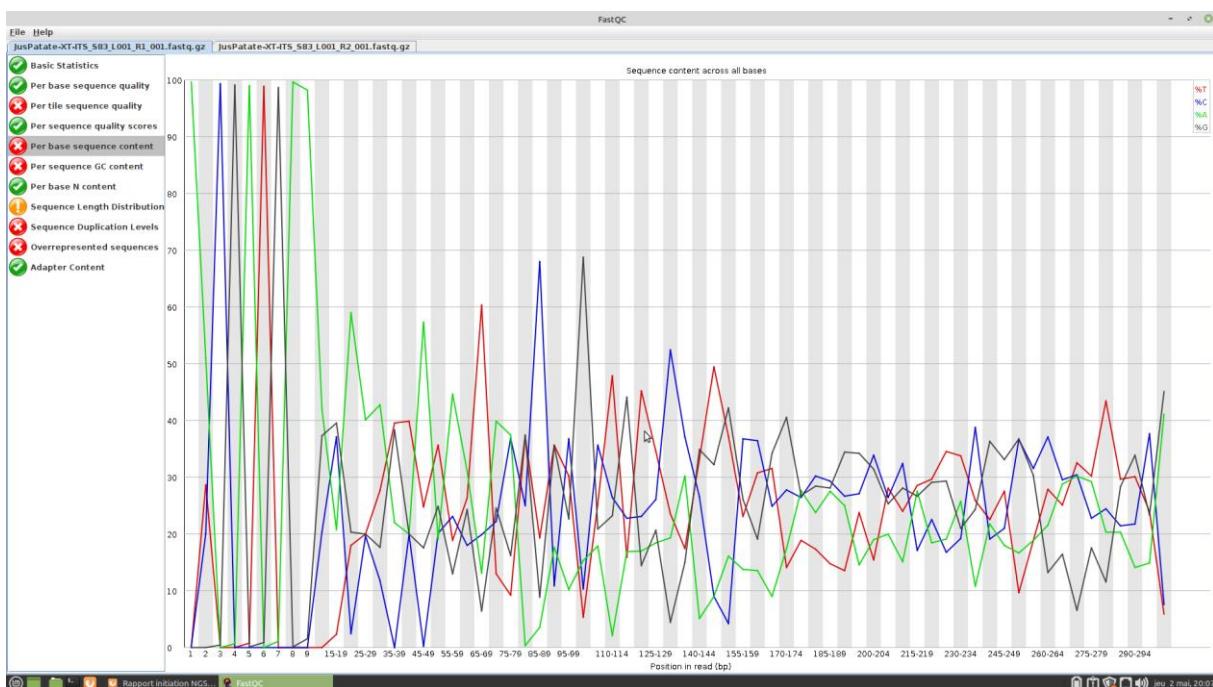


Figure 88 ITS per base sequence content R1 Jus de patate

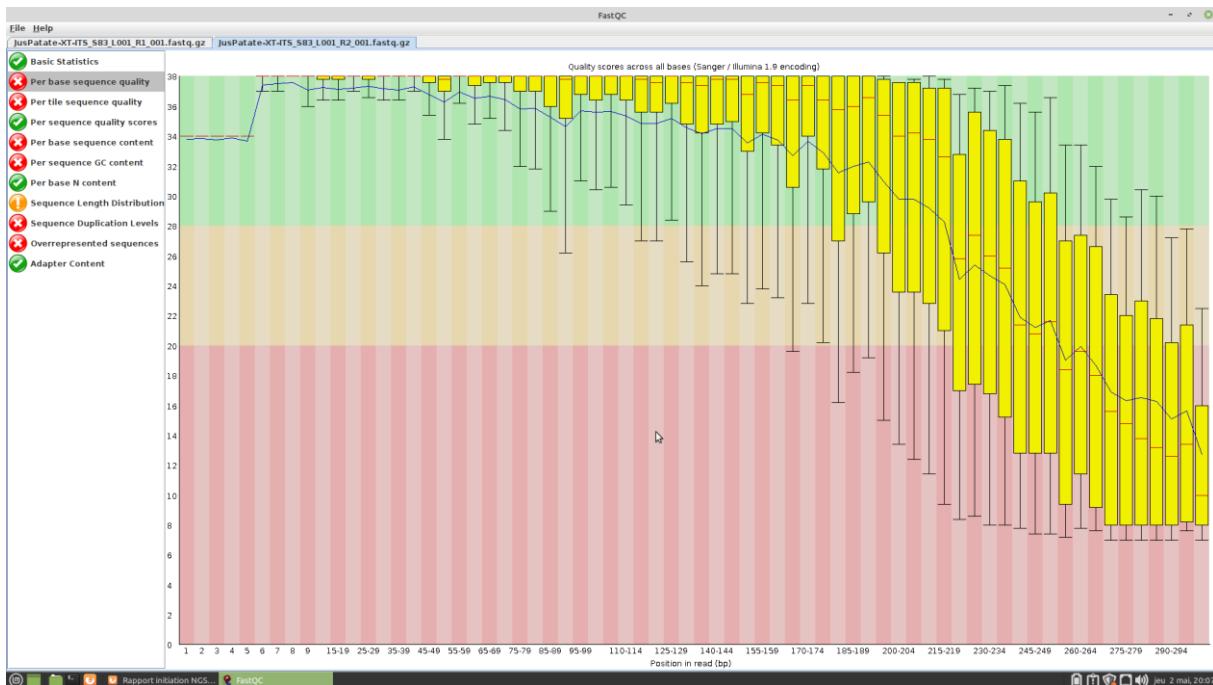


Figure 89 ITS per base sequence quality R2 Jus de patate

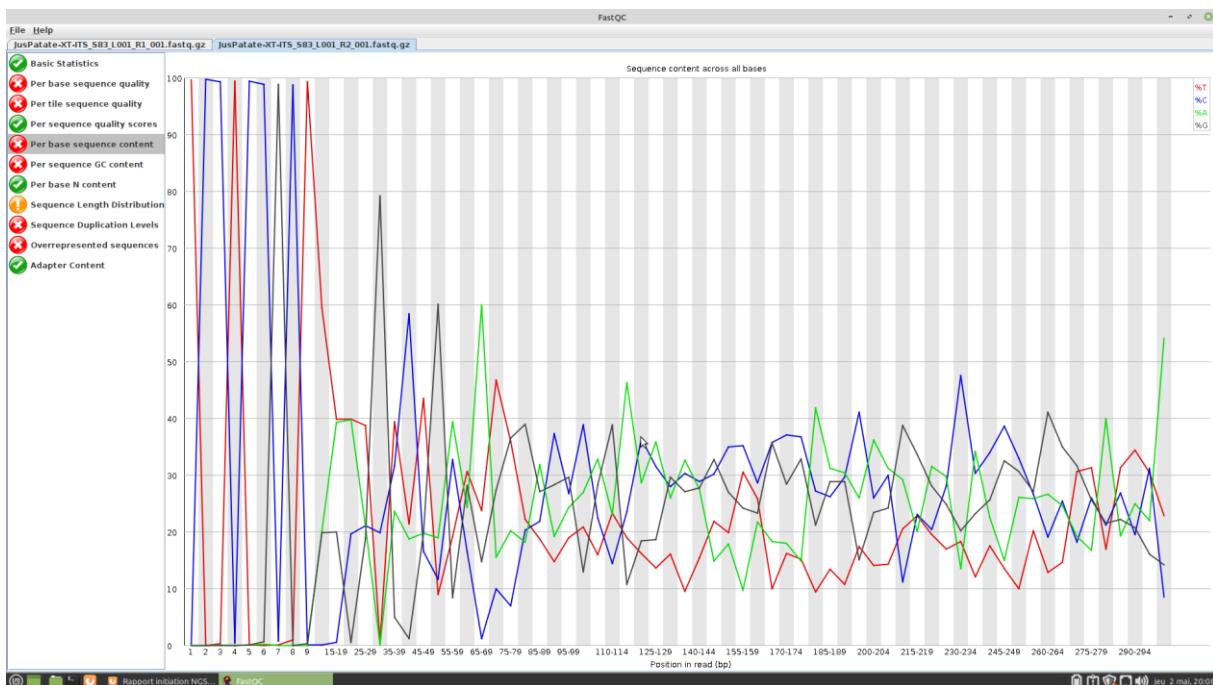


Figure 90 ITS per base sequence content R2 Jus de patate

7.2 Krona

7.2.1 Compost

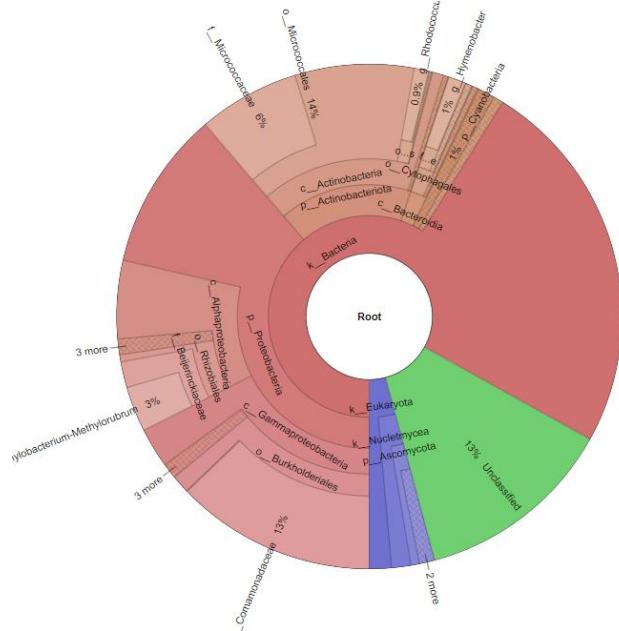


Figure 91 Krona 16S du compost

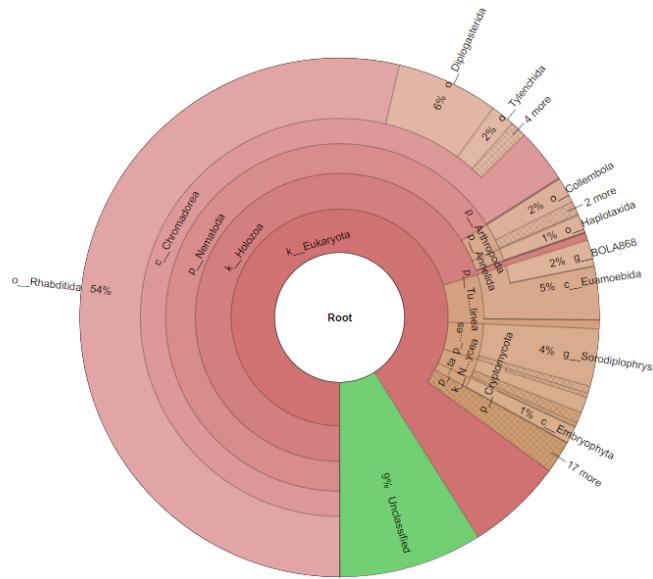


Figure 92 Krona 18S du compost

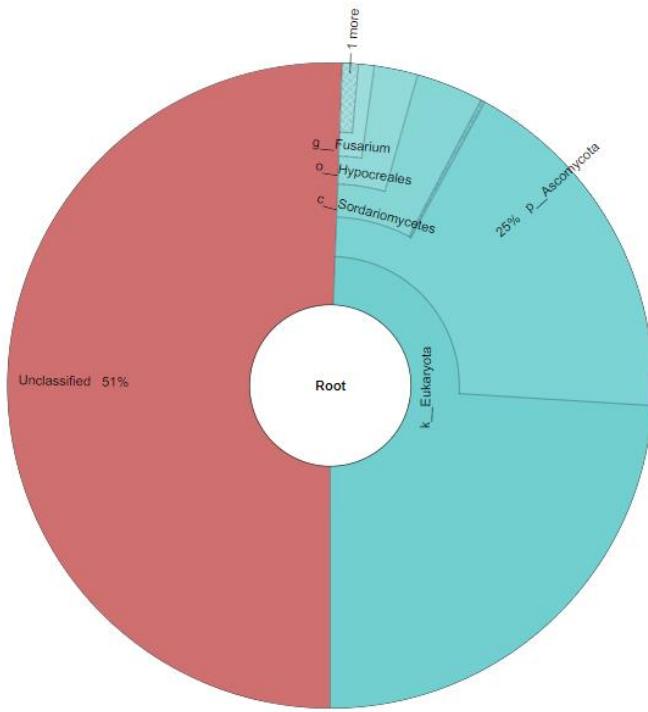


Figure 93 Krona ITS du compost

7.2.2 Eau

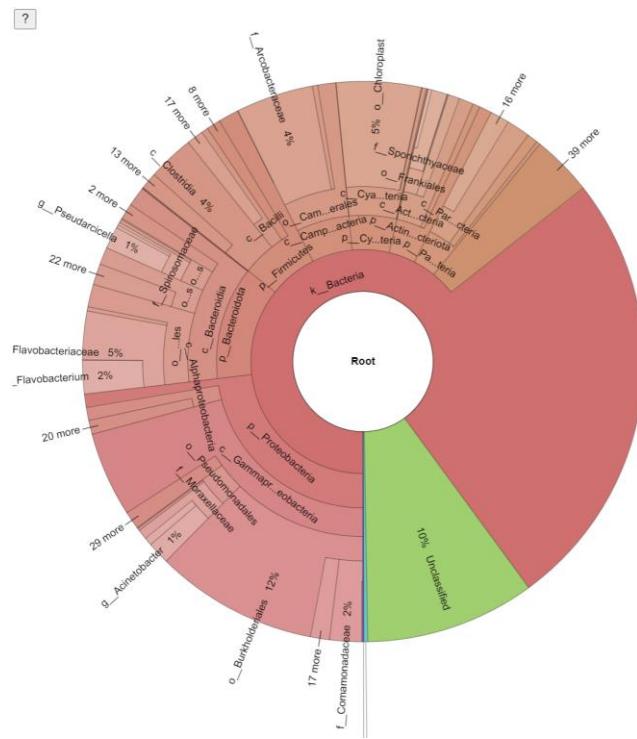


Figure 94 : Représentation Krona eau 165

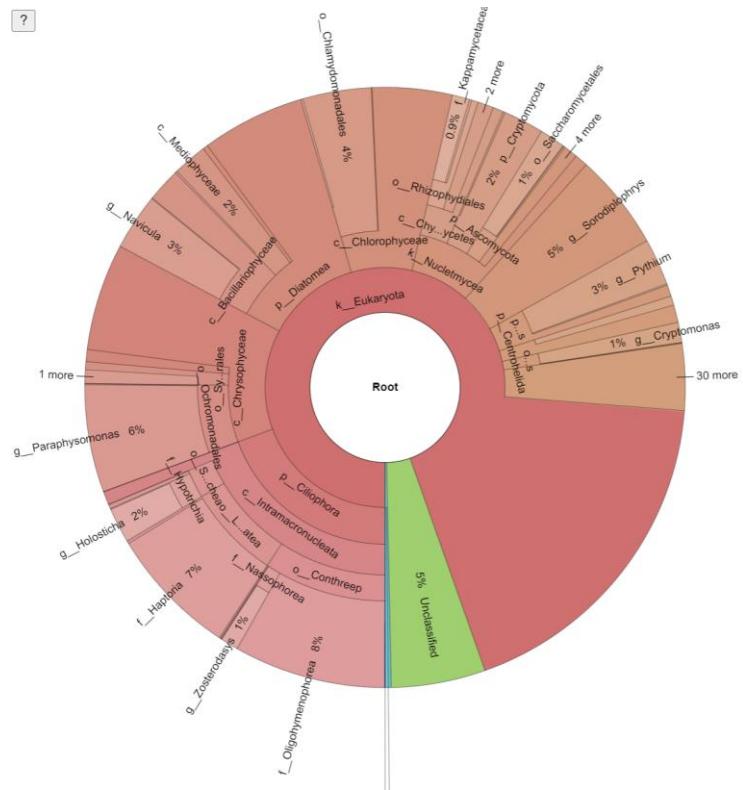


Figure 95 : Représentation Krona eau 18S

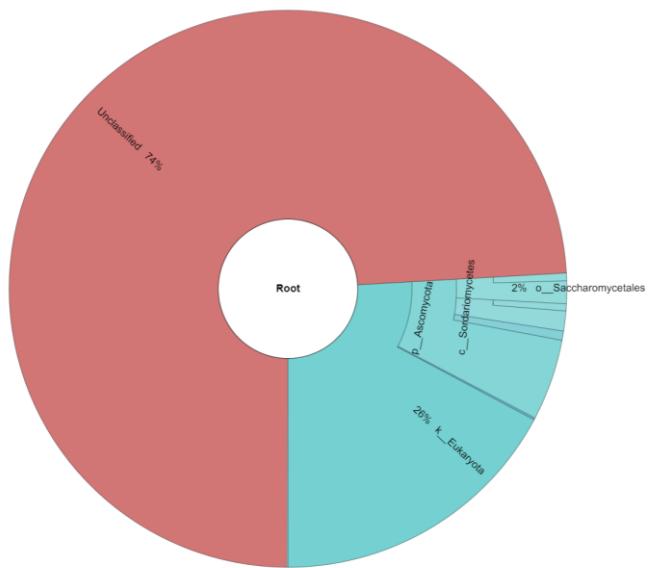


Figure 96 : Représentation Krona Eau ITS

7.2.3 Terre

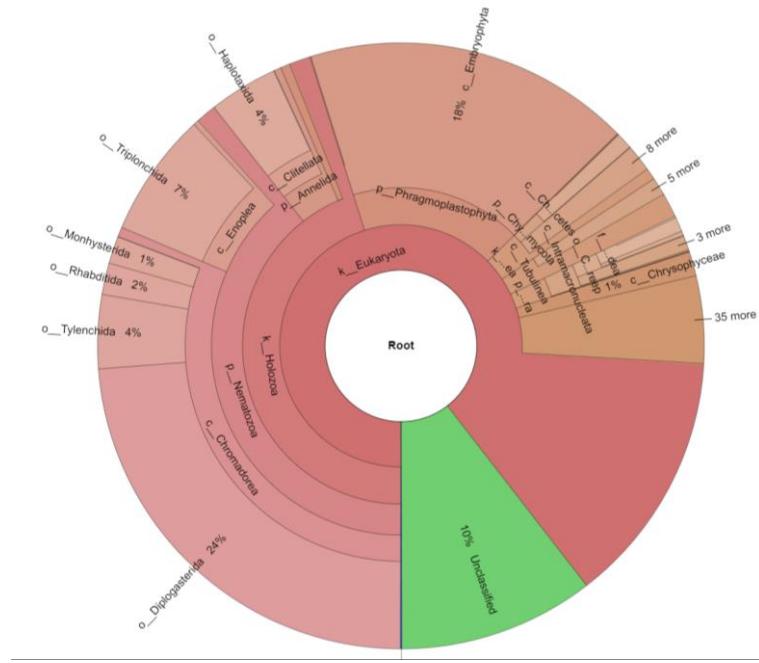


Figure 97 : Représentation Krona Sol 18S

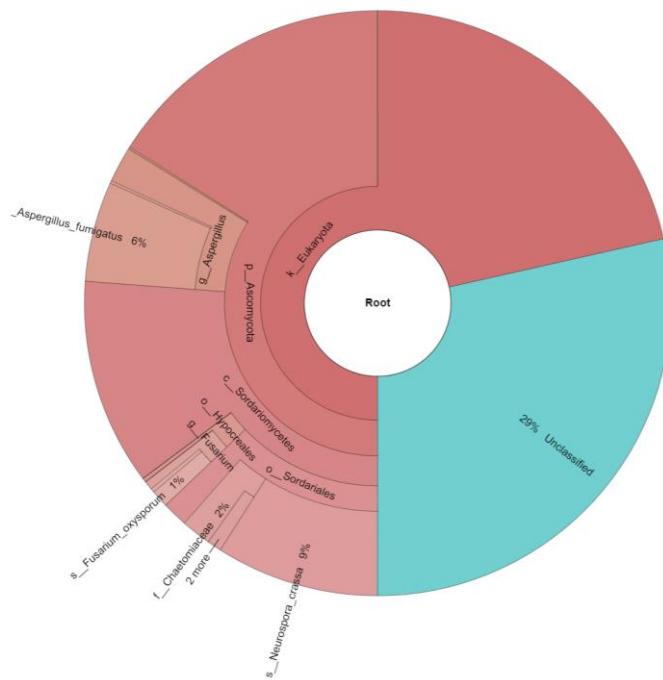


Figure 98 : Représentation krona Sol ITS

7.2.4 Jus de patate

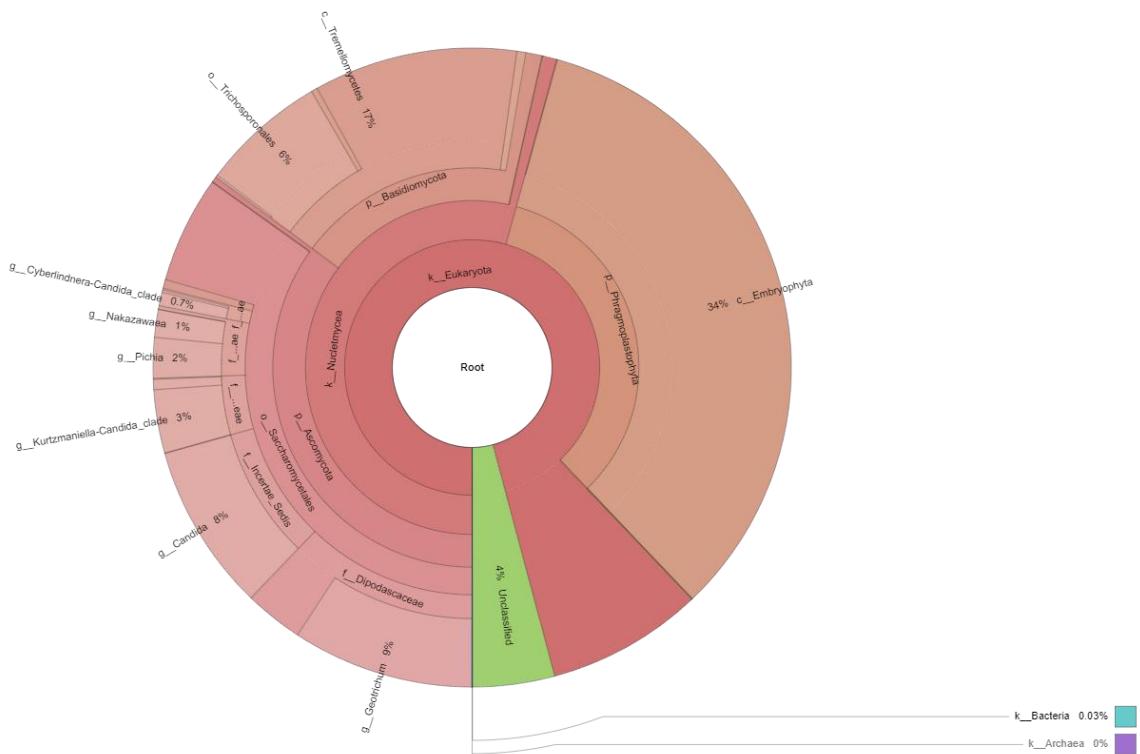


Figure 99 Représentation Krona du 18S Jus patate

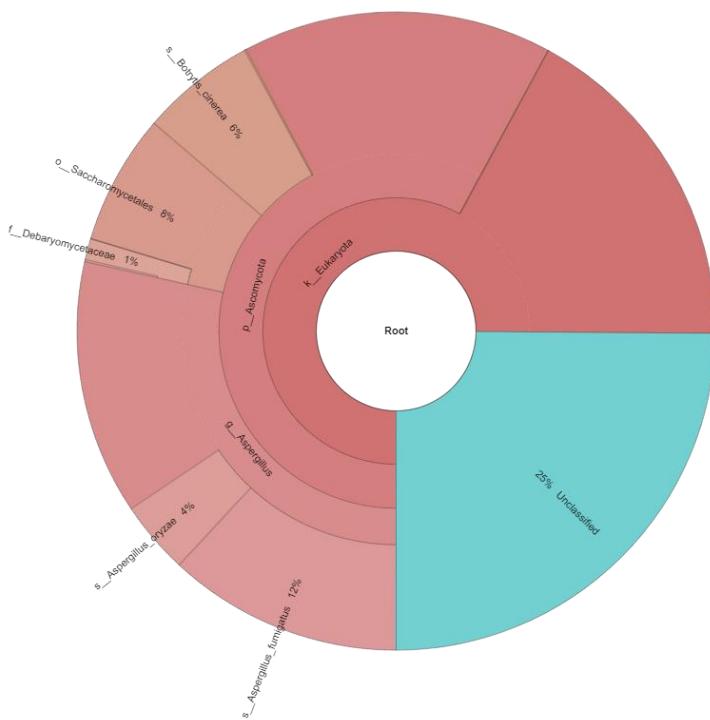


Figure 100 Représentation Krona de l'ITS Jus patate

7.3 Pavian

7.3.1 Composte

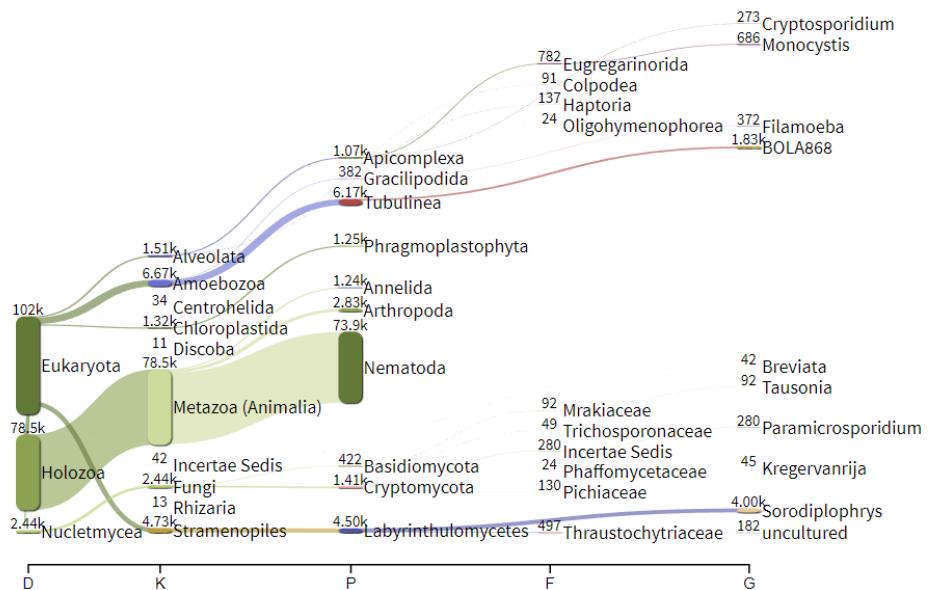


Figure 101 Représentation pavian des 18S du compost

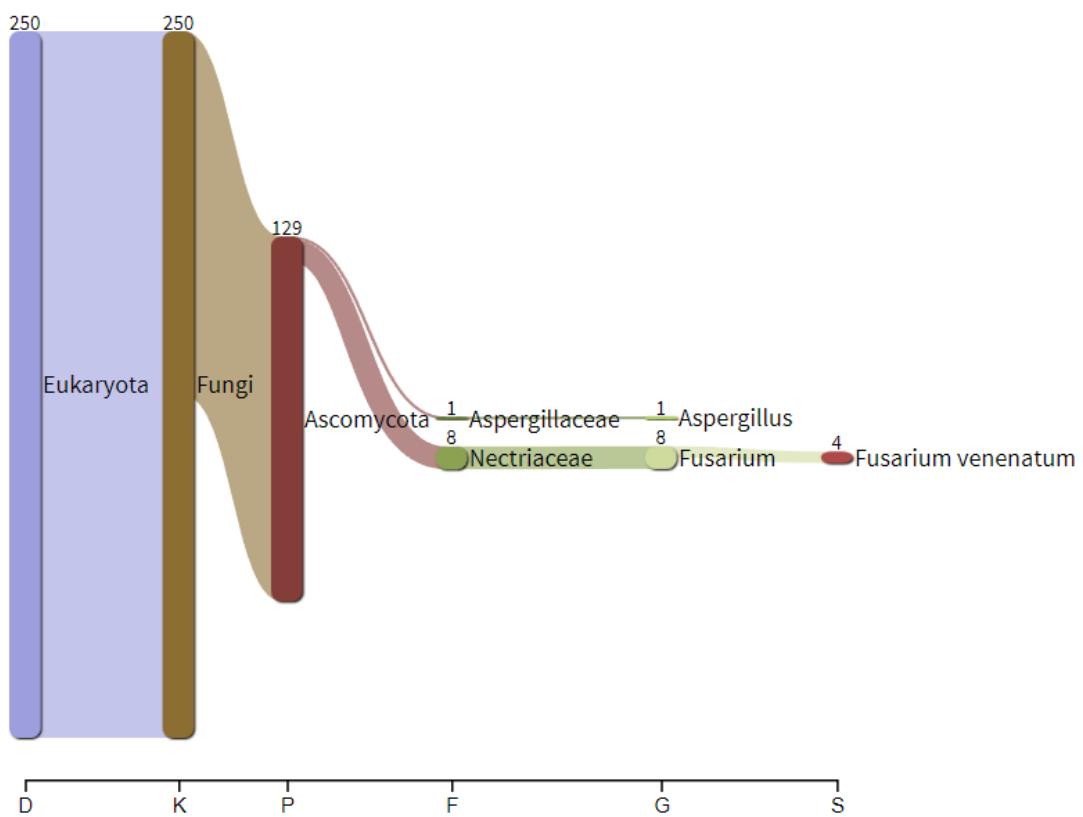


Figure 102 Représentation pavian de l'ITS du compost

7.3.2 Eau

eau16

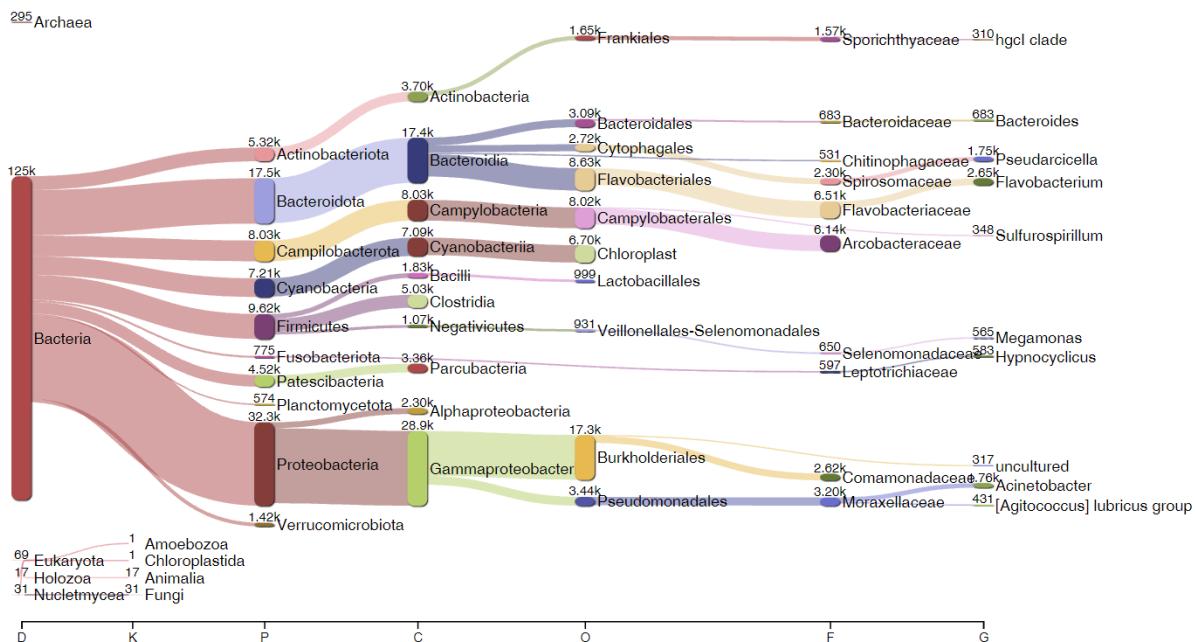


Figure 103 Représentation pavian des 16S de l'eau

eau18

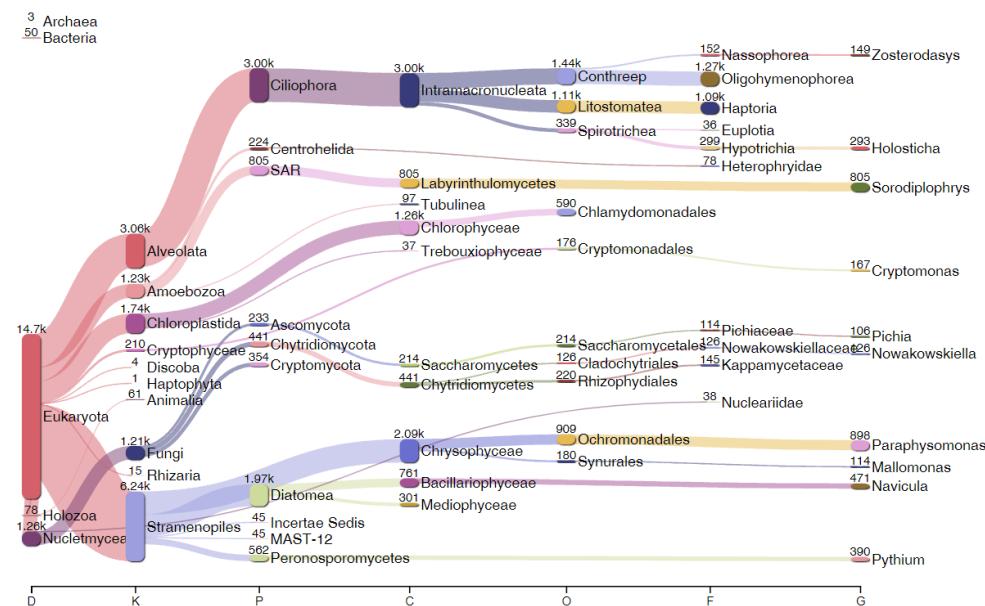


Figure 104 Représentation pavian des 18S de l'eau

eauIT

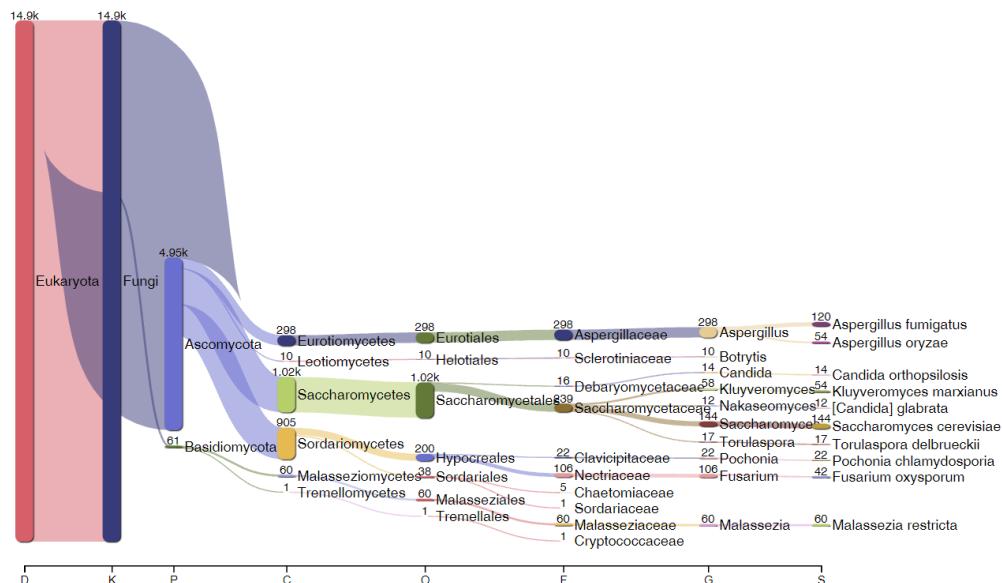


Figure 105 Représentation pavian de l'ITS de l'eau

7.3.3 Terre

sol16

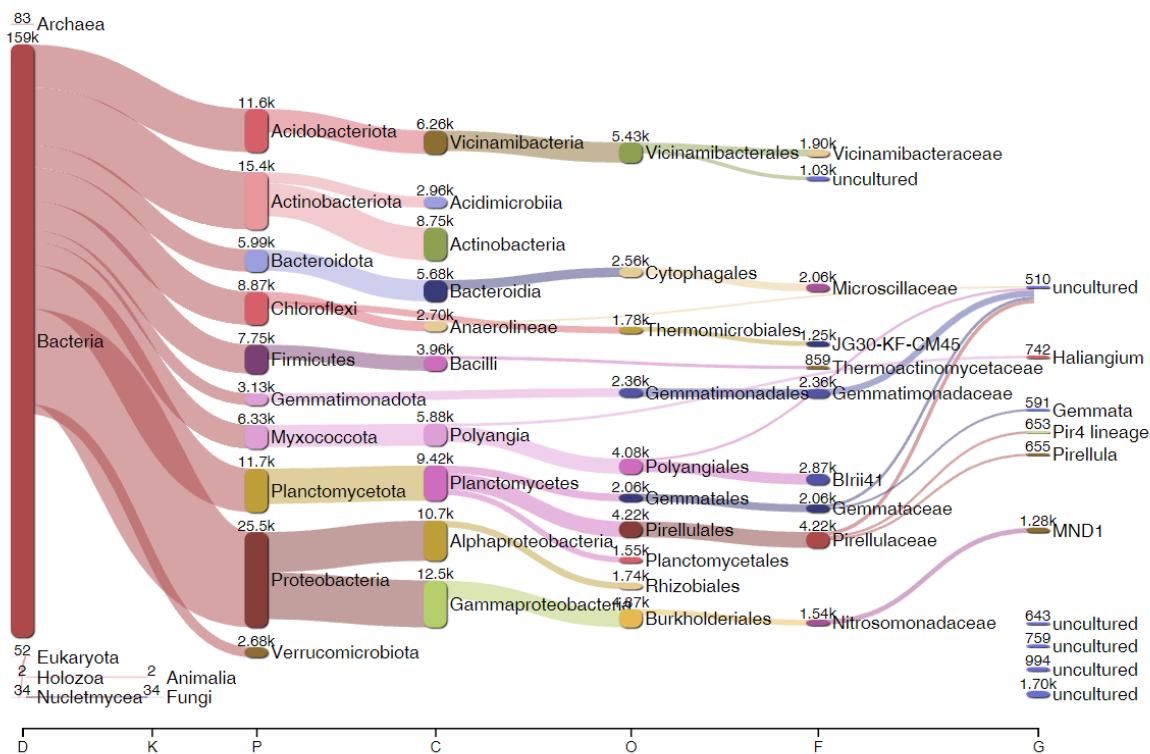


Figure 106 Représentation pavian des 16S du sol

Sol18

60 Bacteria

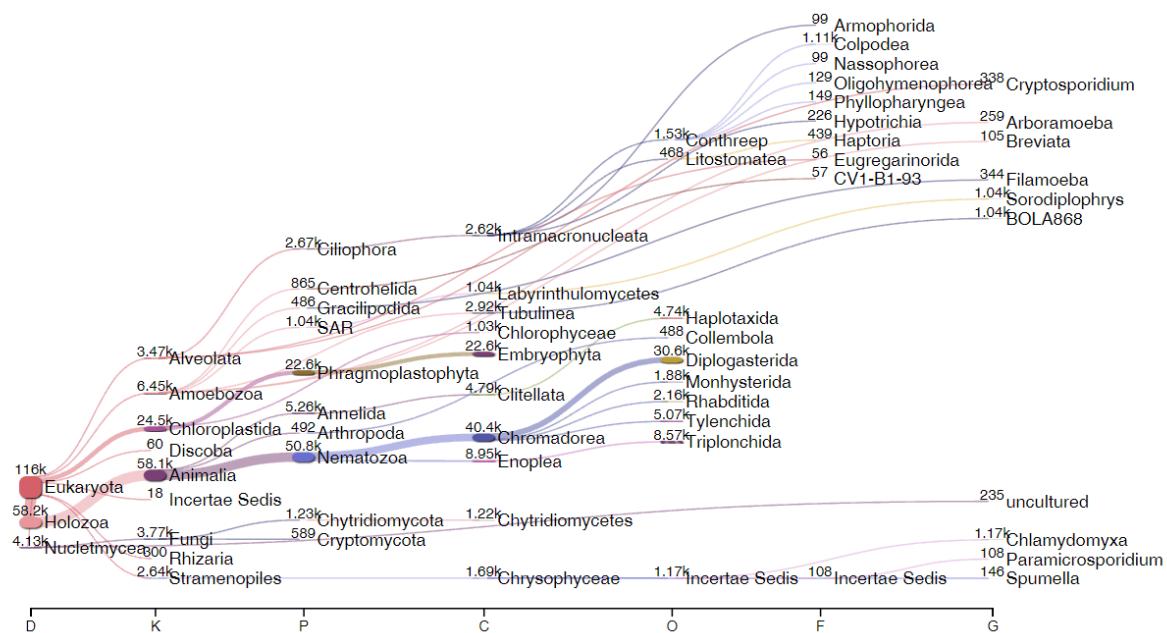


Figure 107 Représentation pavian des 18S du sol

SOlIT

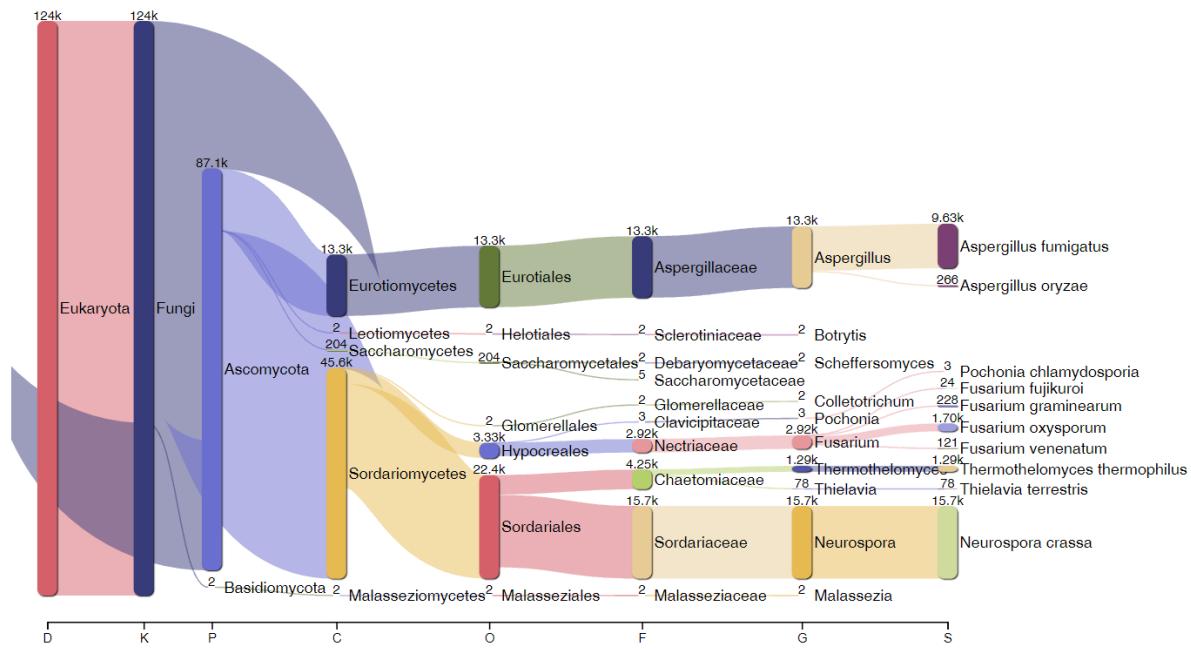


Figure 108 Représentation pavian de l'ITS du sol

7.3.4 Jus de patate

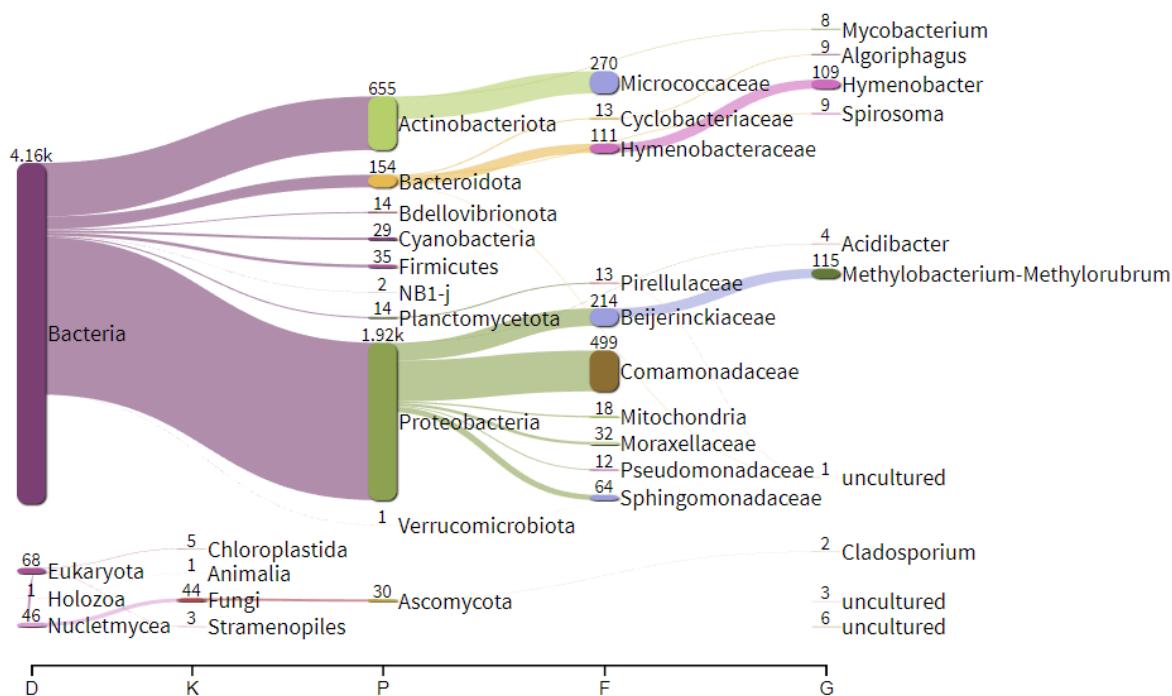


Figure 109 Représentation pavian des 16S du jus de patate

¹ Archaea

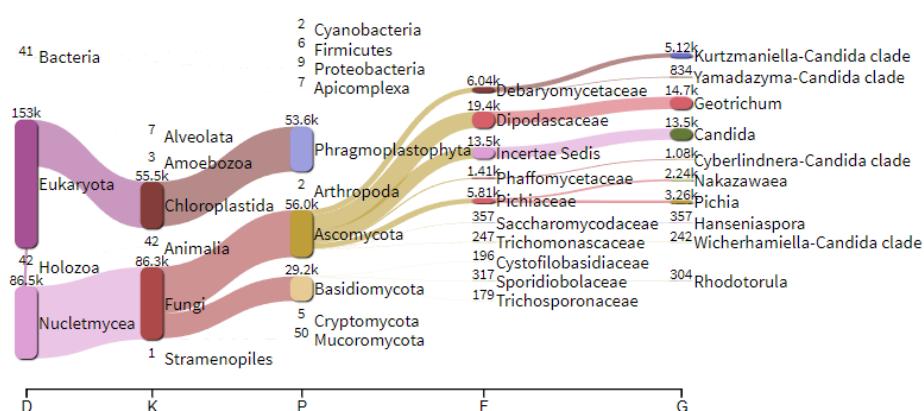


Figure 110 Représentation pavian des 18S du jus de patate

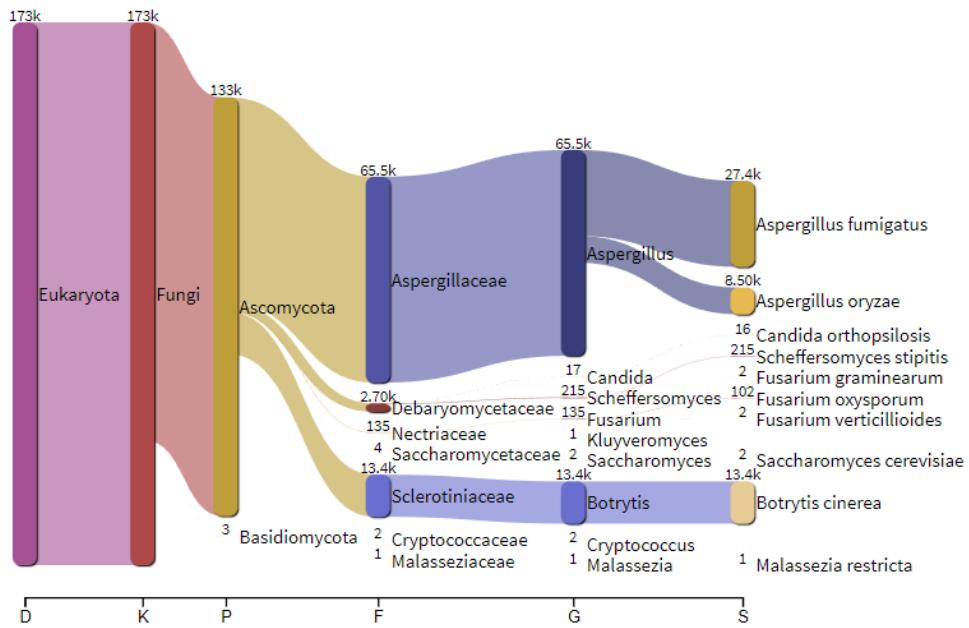


Figure 111 Représentation pavian de l'ITS du jus de patate