Supporting Information Appendix

Doruk Beyter[1], Pei-Zhong Tang[2], Scott Becker[2], Tony Hoang[3], Damla Bilgin[3], Yan Wei Lim[4], Todd C. Peterson[2], Stephen Mayfield[5], Farzad Haerizadeh[2], Jonathan B. Shurin[6], Vineet Bafna[1], and Robert McBride[3]

[1]Computer Science and Engineering Department, University of California San Diego
[2]Life Technologies
[3]Sapphire Energy
[4]Biology Department, San Diego State University
[5]Division of Biological Sciences, University of California San Diego
[6]Section of Ecology, Behavior and Evolution, University of California San Diego

February 3, 2016

Contents

List of Figures

List of Tables

1 Supplementary Methods

1.1 DNA preparation

Each 50 ml biological sample was thawed, homogenized, and two 15ml subsamples withdrawn from the original sample and placed in 15ml tubes. These were centrifuged at 3500rpm for 20 minutes. The supernatant from each sample was combined and transferred to a 50mL tube. This was then concentrated using Amicon Ultra Centrifugal Filters (EMD Milipore, 2015). 15 mL of supernatant was added to Amicon Ultra Centrifugal Filters. These were centrifuged at max (3750rpm) for 1 hour. The liquid was disposed. The remaining supernatant was added to the filter which was again centrifuged at max (3750rpm) for 1 hour. 200μL from the top of the filter was transferred into a new centrifuge tube and stored. This liquid was then added to the pellet from the original centrifuge and DNA extracted using the PowerLyser PowerSoil DNA isolation Kit (Mo Bio Laboratories Inc., 2015).

The DNA from the extraction was amplified using primers designed to target both the V4 region of 16S rRNA gene and the ITS2 region of prokaryotic and eukaryotic genomes. Primers were ordered with with 5' PHO modifications to ensure compatibility with labeling for the sequencing steps. The amplicon for the 16S should fall approximately between the 100-400bp range and the primers were designed to universally target Archea and Bacteria (Forward: S-D-Bact-0564-a-S-15 (41345) AYTGGGYDTAAAGNG, Reverse: S-D-Bact-0785-b-A-18 (41346) TACNVGGGTATCTAATCC). The amplicon for the ITS2 primer should fall approximately between 200-400bp and were selected because they universally target eukaryotes (Forward: (41343) GCATCGATGAAGAACGCAGC, Reverse: (41344) TCCTCCGCTTATTGATATGC).

The PCR was set up in a 96 well plate as follows: 20.0μL 5X HF buffer (Phusion kit), 4.0μL 10 mM dNTPs (NEB), 4.0μL DMSO (Phusion kit), 10.0μL 5M Betaine, 5.0μL 10μM of each primer, 0.8μL Phusion polymerase, 6.0μL DNA template. To cover the diversity represented gradient PCR was performed with the following PCR protocol: 98°C 0:30, 25X (98°C 0:10, 43°C-53°C 0:30, 72°C 0:30), 72°C 5:00, 4°C hold. Gels were run to ensure correct band sizes. The DNA was then pooled and cleaned using Invitrogen PureLink Pro 96 PCR purification Kit (Life Technologies, 2015). The resultant DNA was then quantified to ensure 2 micrograms and prepped for sequencing.

1.2 TMAP usage

We applied the "map2" algorithm (based off of the BWA long-read algorithm [4]), designed for reads longer than 150bps, due to the read sizes (a mean of 240bps for 16S and 420 for ITS2 sequences – see Figures S4, S5, and S6 for read length distributions in all chips and samples; individually, and all combined) and other default parameters associated with it. For every read, TMAP returns the mapping with the best score. If multiple sequences had the same best score, a random mapping among them was returned.

1.3 OTU-based analysis for 16S data

Several OTU-based pipelines such as UPARSE [5], QIIME [6], MOTHUR [7] have been developed for the analysis of Illumina or 454 pyrosequencing 16S and fungal only ITS2 marker-gene sequencing data. Very recently, a pipeline that includes 16S Ion Torrent PGM sequencing is developed [8], and used it in the Brazilian Microbiome Project (BMP) [9]. The BMP 16S profiling analysis pipeline makes use of the UPARSE OTU clustering, and QIIME taxonomy assignment, using Ribosomal Database Project (RDP) naive classifier [10].

In order to compare our 16S data analysis results with OTU-based pipelines, we used the pipeline suggested by BMP. We began by truncating the reads at length 200 as the read ends are assumed to have lowered quality, and discarded any read with a smaller length. We then removed any read having an expected error rate of 1.0, a suggested value in the UPARSE documentation [11]. We applied dereplication that removes the identical reads for faster querying, and removed any singleton reads. We clustered the OTUs, and applied a reference based chimera filtering using a gold database, which contains the ChimeraSlayer reference database from the Broad Microbiome Utilities version microbiomeutil-r20110519, as described in [12], using the plus strand, as specified. We finally assigned all quality filtered reads, including the singletons, to the constructed OTUs at 97% identity. All analysis until this point was performed using usearch v7.0.1090_i86linux32. We gathered the taxonomy information using assign_taxonomy.py version 1.7.0 from QIIME, choosing RDP classifier as taxonomy assignment algorithm with the default bootstrap confidence threshold of 80%, and OTUs pre-constructed from GreenGenes (version May 2013) at 97% identity, as training sequences.

1.4 Comparison of sequence mapping and OTU-based approaches and reproducibility assessment among chips

We performed a Mantel test between the sample taxonomy composition results of our approach and the BMP pipeline for 16S data analysis as follows: at ranks phylum, class, order, family and genus, respectively we obtained the taxonomies of both analysis results. We took the union of the taxonomies observed in the two analyses, and assigned abundance values of 0 to any taxonomy in the union set not observed in individual results, for all 26 time point samples. Thus, for each approach, we had pairs of relative abundance values for all taxonomies in the union set at all time points as a matrix, which we called *a taxonomy abundance matrix*, for each of the aforementioned rank. We compared these pairs of taxonomy abundance matrices using the package "ade4" [13] in R with the function "mantel.rtest" using 999 replicates. We achieved Mantel r statistics of 0.99, 0.98, 0.94, 0.94, 0.91 for ranks phylum, class, order, family, and genus, respectively, all with p-value 0.001, suggesting high result similarity. Since the RDP classifier is not capable in classification beyond the genus level, we have no comparison available with the BMP pipeline at species/sequence level of resolution. BMP pipeline area plots at ranks phylum, class, and genus are shown in Figure S14, for visual comparison purposes.

We also note that a 16S genus level diversity comparison between the two approaches yield a nearly identical pattern: the linear regression describing the relationship between the two was: $r^2 = 0.96$, $P = 2.60 \cdot 10^{-14}$.

The reproducibility assessment among chips for 16S and ITS2 data also follows the same Mantel test approach, with the single difference of containing the top 2000 and 200 sequence relative abundances (instead of taxa relative abundances) in the compared pairs of abundance matrices coming from different chips.

1.5 Challenges in OTU-based approaches and taxonomy assignment on ITS2 data

Given the high variance in the ITS2 region length, ranging from 100bps to 700bps [14]; length trimming, a critically important step in an OTU-based approach [11], is not practical. Moreover, the taxon dependent OTU clustering identity percentages on microbial eukaryotes [15], may render the OTU clustering step erroneous. The taxon dependency of OTU clustering identity percentages also makes the RDP naive

Bayesian classifier taxonomy assignment (used in OTU-based approach) challenging, as its reference taxonomy database is expected to be clustered at a certain identity percentage. Another challenge in contructing a clustered ITS2 database from NCBI would lie in determining the correct boundaries of the ITS2 region, previous to clustering, due to the flanking 18S, ITS1, 5.8S, and 28S regions in the NCBI nucleotide entries. Previous research [16] reports that taxonomy classification results using BLASTN, a mapping based approach, and RDP naive Bayesian classifier are very similar on ITS2 data. Considering these challenges and findings, we preferred to determine the taxa relative abundances using a mapping approach.

1.6 Outlier removal on time series ecosystem data

We initially subtracted the 7-day local central mean from each data point. We perfomed this step in order to reduce the dependency between successive points in our time series ecosystem data and to satisfy the idenpendent, identicaly distribution requirement for a normal distribution. We, then, tested for normality using "shapiro.test" in R, using the package "stats" [17]. Upon confirming for normality, we removed any data point that exceeded 3σ of distance from mean. We did not perform outlier detection for NH_4, urea, NO_3, NO_2, and PO_4, due to the expected high fluctuations stemming from pond nutrient management.

1.7 Model comparison using F-test

In order to explore the explanatory values of certain factors on a target, controlling for other factor(s), we compared two models: a reduced and a full model. The reduced model contains the factor we would like to control for, whereas the full model contains additional factor(s), which we are interested to explore the effect on our target.

$$\begin{aligned} \text{Reduced Model} \quad y =& \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon_r \\ \text{Full Model} \quad y =& \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \beta_{k+1} x_{k+1} + \cdots + \beta_p x_p + \varepsilon_f \end{aligned} \tag{1}$$

where in one our tests, for instance, y was chosen as the eukaryotic diversity we were targeting, $x_1, \ldots, x_k$ as the factors we controlled for such as temperature and bacteria diversity, and $x_{k+1}, \ldots, x_p$ as any factor(s) we explored the effect it had on the target, such as pre- and post-pesticide sampling. We tested if we could reject the null hypothesis:

$$H_0 : \beta_{k+1} = \cdots = \beta_p = 0$$

to see if our full model added a significant explanatory value over the reduced model, using an F statistic:

$$F = \frac{(RSS_{reduced} - RSS_{full})/(p-k)}{RSS_{full}/(n-p-1)} \tag{2}$$

where RSS_i is the residual sum of squares of model i.

2 Supplementary Results

2.1 Mapping statistics

We initially discarded any read having length shorter than 50 nucleotides, and an error rate higher than 2.0 for 16S reads, and 4.0 for ITS reads, due to their longer average size compared to 16S. After mapping the remaining 16S and ITS2 reads to respective databases, we calculated percent identity, and *query-coverage*, defined as the fraction of the query sequence matching to the target, for assessing mapping quality. For these measures, the quality was uniformly high with a mean percent identity of 97% and 96%, and mean coverage over 94% and 82% across all 16S and ITS2 reads that mapped their respective database. (Figures S7 and S8). Following the cutoffs applied by "16S Ribosomal RNA Reference Sequence Similarity Search" by NCBI [18], we used a 95% percent identity and 70% of query-coverage cutoff. On average among all chips, 75% of the 16S and 77% of the ITS2 reads exceeded our chosen cut-offs, and were used in subsequent analyses.

2.2 Intra-sample reproducibility assessment

In order to assess robustness in the sample composition analyses, two redundant samples were used as technical replicates for each of samples 4, 11, 19 and 24, in the design (samples 27 and 31 were replicates of sample 4, 28 and 32 for 11, 29 and 33 for 19, and 30 and 34 for 24). Figure S9 demonstrates that the technical replicates consistently show low dissimilarity values (mean Bray Curtis dissimilarity values of 0.06, 0.03, 0.04, 0.02 and 0.04, 0.07, 0.50, 0.06, for the two replicates of samples 4, 11, 19 and 24 for 16S and ITS2, chip 3.) suggesting good reproducibility, except sample 19 for ITS2 data only. We note the replicates for sample 19 (samples 29 and 33, ITS2 data) had a skewed read length distribution, compared to sample 19 itself, (see Figure S5b), which might be a possible reason for the observed noise.

2.3 Pre- and post-fungicide relationship of productivity variability and temperature

We investigated whether temperature, based on its pre-fungicide era relationship with productivity variability (standard deviation), could predict the post-fungicide productivity standard deviation (sd) trends. Figure S17 shows linear relationship between temperature and productivity sd in different periods. During the pre-fungicide period, temperature showed a positive correlation with productivity sd, whereas it had a negative correlation during the post-fungicide period, therefore temperature alone cannot explain the change in the productivity variability observed after the fungicide application.

References

[1] EMD Milipore. UFC903008 — Amicon Ultra-15 Centrifugal Filter Unit with Ultracel-30 membrane. Available at: `http://www.millipore.com/catalogue/item/ufc903008`, 2015. Last Accessed: 01 April 2015.

[2] Mo Bio Laboratories Inc. Powerlyzer Powersoil DNA Isolation Kit. Available at: `http://www.mobio.com/soil-dna-isolation/powerlyzer-powersoil-dna-isolation-kit.html`, 2015. Last Accessed: 01 April 2015.

[3] Life Technologies. Purelink Pro 96 PCR Purification Kit. Available at: `http://products.invitrogen.com/ivgn/product/K310096A`, 2015. Last Accessed: 01 April 2015.

[4] Heng Li and Richard Durbin. Fast and accurate long-read alignment with burrows–wheeler transform. *Bioinformatics*, 26(5):589–595, 2010.

[5] Robert C Edgar. Uparse: highly accurate otu sequences from microbial amplicon reads. *Nature methods*, 10(10):996–998, 2013.

[6] J Gregory Caporaso, Justin Kuczynski, Jesse Stombaugh, Kyle Bittinger, Frederic D Bushman, Elizabeth K Costello, Noah Fierer, Antonio Gonzalez Pena, Julia K Goodrich, Jeffrey I Gordon, et al. Qiime allows analysis of high-throughput community sequencing data. *Nature methods*, 7(5):335–336, 2010.

[7] Patrick D Schloss, Sarah L Westcott, Thomas Ryabin, Justine R Hall, Martin Hartmann, Emily B Hollister, Ryan A Lesniewski, Brian B Oakley, Donovan H Parks, Courtney J Robinson, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology*, 75(23):7537–7541, 2009.

[8] Victor S Pylro, Luiz Fernando W Roesch, Daniel K Morais, Ian M Clark, Penny R Hirsch, and Marcos R Tótola. Data analysis for 16s microbial profiling from different benchtop sequencing platforms. *Journal of microbiological methods*, 107:30–37, 2014.

[9] Victor Satler Pylro. Brazilian Microbiome Project. Available at: `http://www.brmicrobiome.org`, 2015. Last Accessed: 01 April 2015.

[10] Qiong Wang, George M Garrity, James M Tiedje, and James R Cole. Naive bayesian classifier for rapid assignment of rrna sequences into the new bacterial taxonomy. *Applied and environmental microbiology*, 73(16):5261–5267, 2007.

[11] Robert C Edgar. UPARSE Pipeline. Available at: `http://drive5.com/usearch/manual/uparse_pipeline.html`, 2015. Last Accessed: 01 April 2015.

[12] Robert C Edgar. UCHIME. Available at: `http://drive5.com/uchime/uchime_download.html`, 2015. Last Accessed: 01 April 2015.

[13] S. Dray and A.B. Dufour. The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22(4):1–20, 2007.

[14] Hui Yao, Jingyuan Song, Chang Liu, Kun Luo, Jianping Han, Ying Li, Xiaohui Pang, Hongxi Xu, Yingjie Zhu, Peigen Xiao, et al. Use of its2 region as the universal dna barcode for plants and animals. *PloS one*, 5(10):e13102, 2010.

[15] Jean-David Grattepanche, Luciana F Santoferrara, George B McManus, and Laura A Katz. Diversity of diversity: conceptual and methodological differences in biodiversity estimates of eukaryotic microbes as compared to bacteria. *Trends in microbiology*, 22(8):432–437, 2014.

[16] Andrea Porras-Alfaro, Kuan-Liang Liu, Cheryl R Kuske, and Gary Xie. From genus to phylum: large-subunit and internal transcribed spacer rrna operon regions show similar classification accuracies influenced by database composition. *Applied and environmental microbiology*, 80(3):829–840, 2014.

[17] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2014.

[18] NCBI. Available at: `http://www.ncbi.nlm.nih.gov/genomes/16S/help.html#query`, 2015. Last Accessed: 12 July 2015.

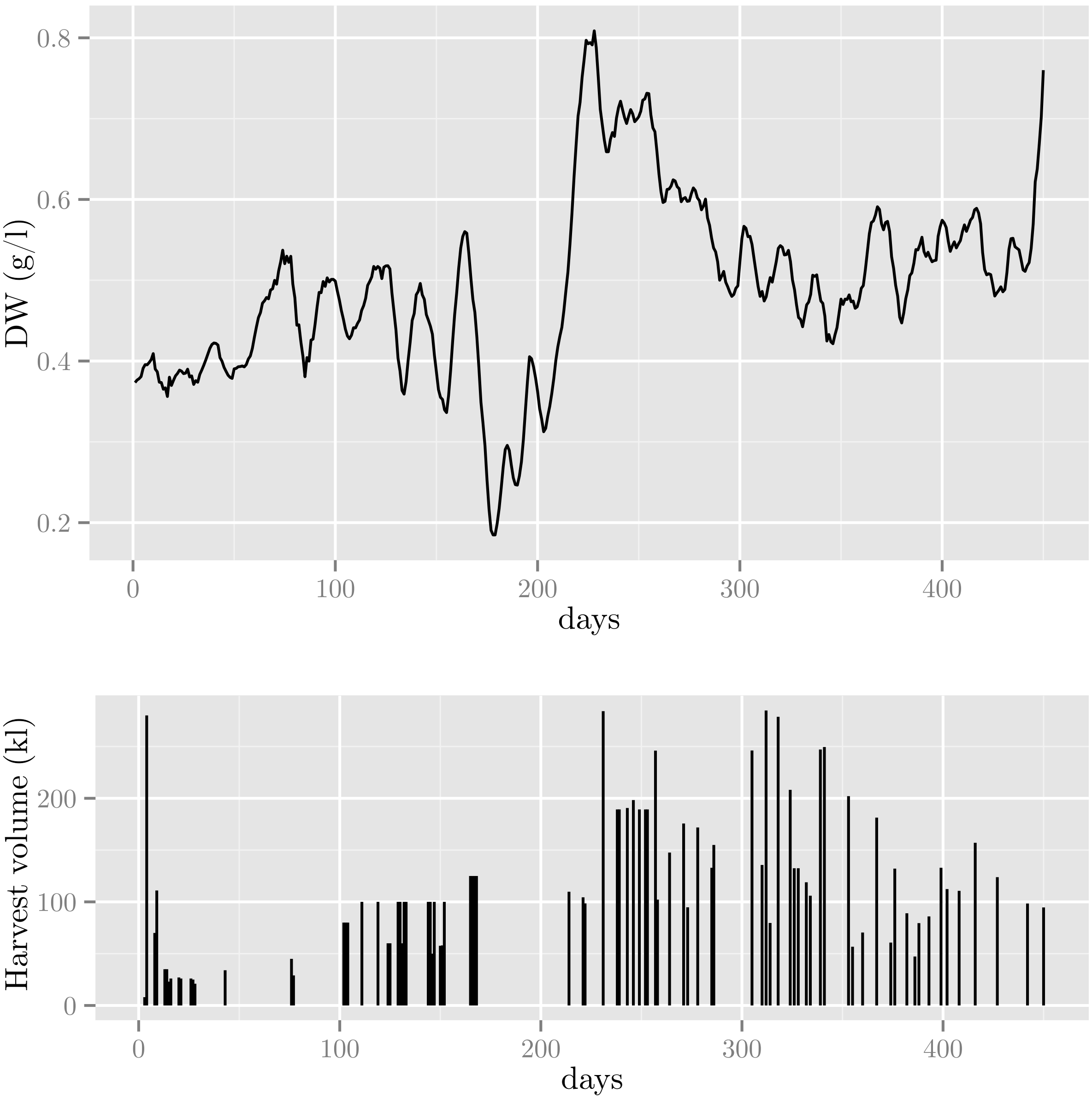


Figure 1: DW (g/l) and harvest volume (kl) in time.

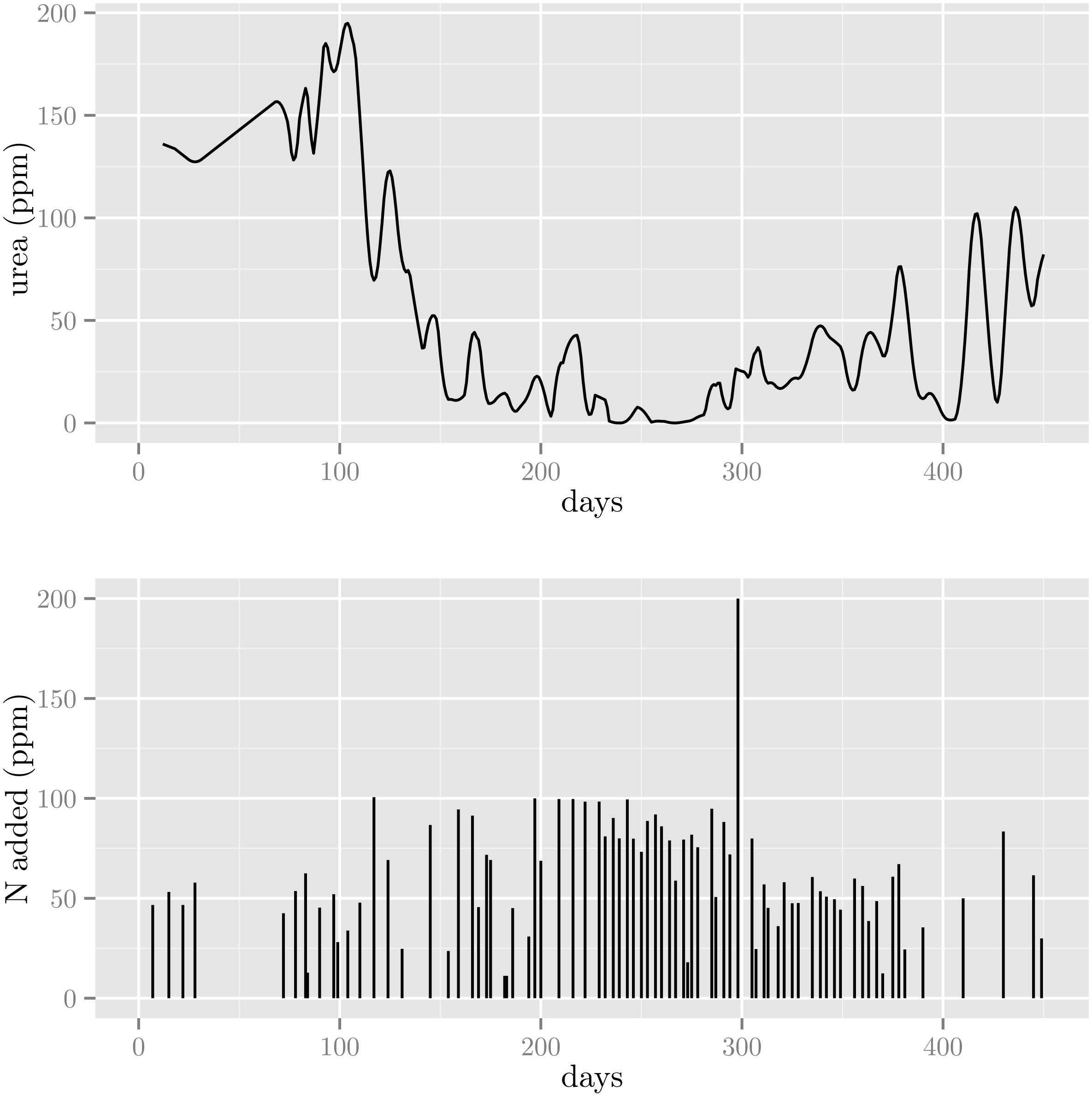


Figure 2: Measured urea levels and N addition (mostly through urea addition) data.

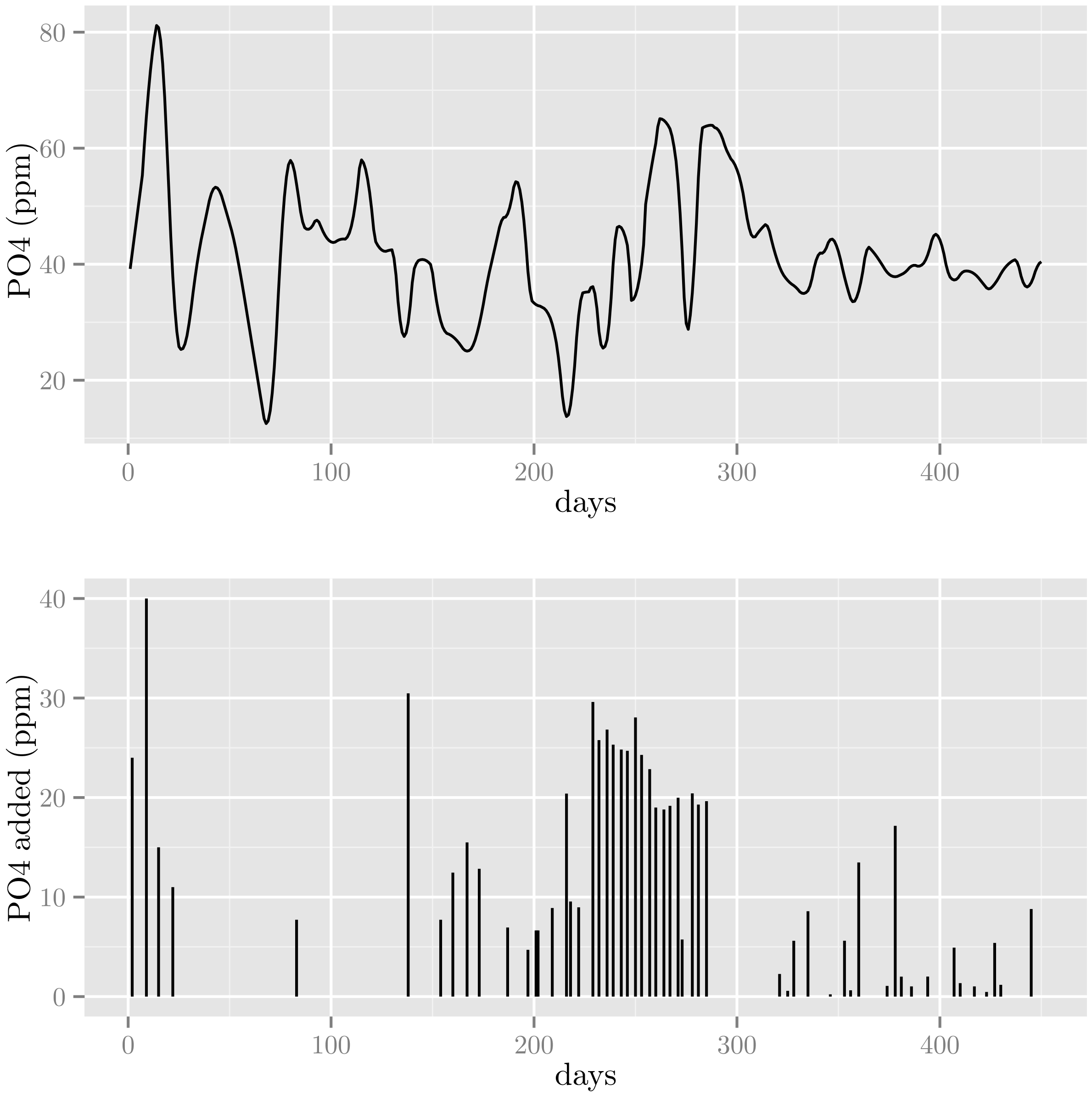


Figure 3: Measured PO4 levels and PO4 addition data.

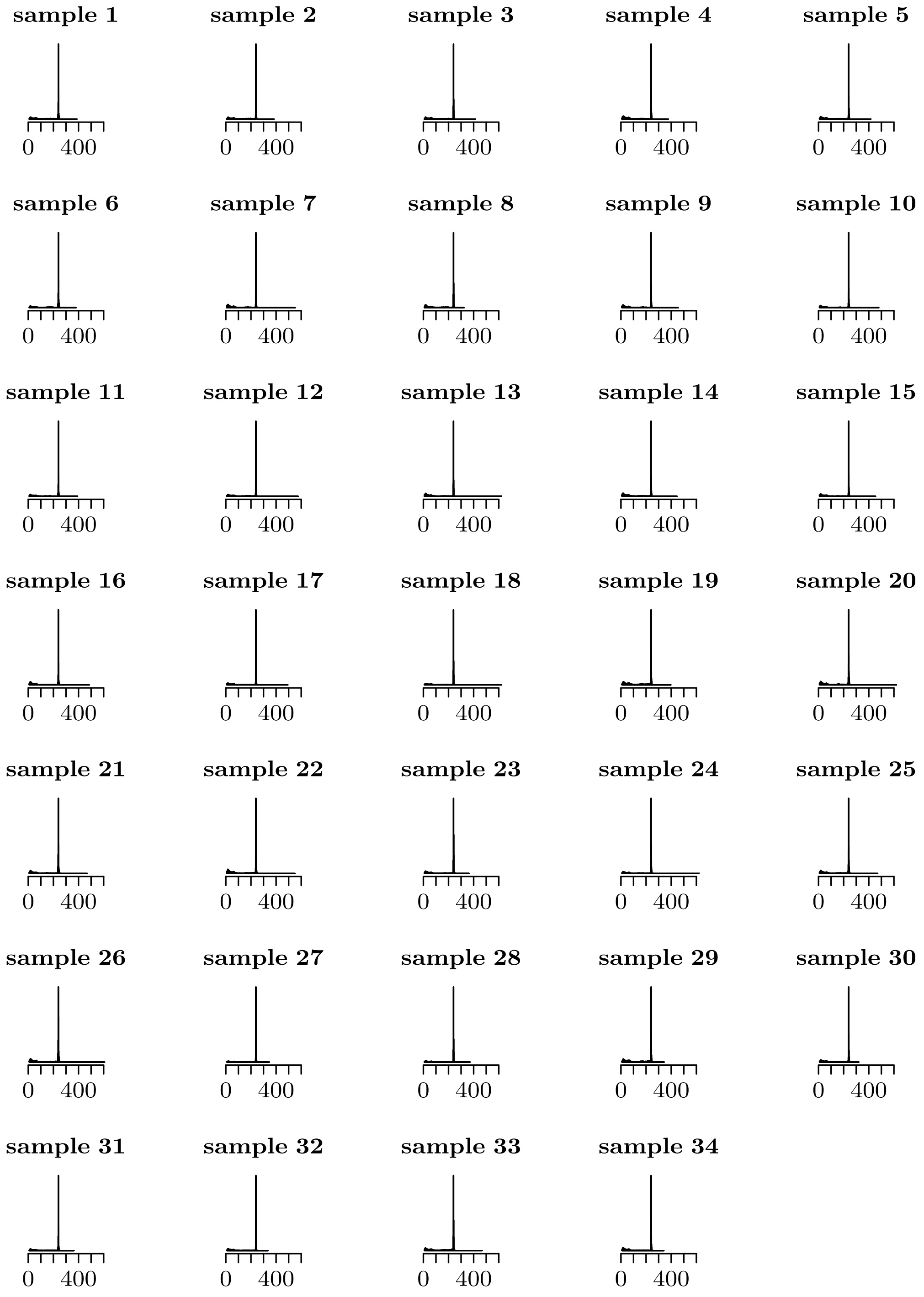

sample 1
0
400
sample 2
0
400
sample 3
0
400
sample 4
0
400
sample 5
0
400
sample 6
0
400
sample 7
0
400
sample 8
0
400
sample 9
0
400
sample 10
0
400
sample 11
0
400
sample 12
0
400
sample 13
0
400
sample 14
0
400
sample 15
0
400
sample 16
0
400
sample 17
0
400
sample 18
0
400
sample 19
0
400
sample 20
0
400
sample 21
0
400
sample 22
0
400
sample 23
0
400
sample 24
0
400
sample 25
0
400
sample 26
0
400
sample 27
0
400
sample 28
0
400
sample 29
0
400
sample 30
0
400
sample 31
0
400
sample 32
0
400
sample 33
0
400
sample 34
0
400

Figure 4a

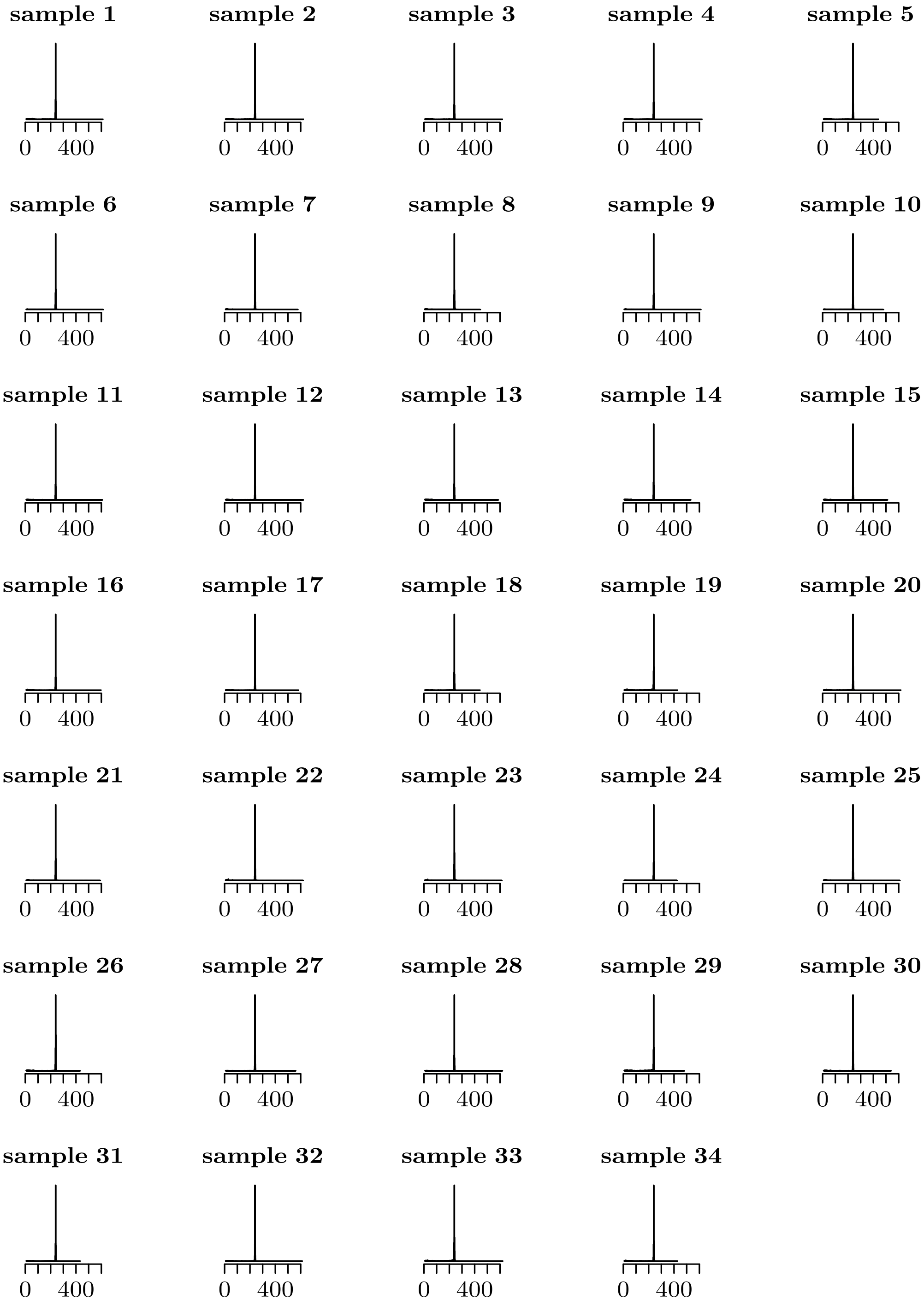

sample 1
0
400
sample 2
0
400
sample 3
0
400
sample 4
0
400
sample 5
0
400
sample 6
0
400
sample 7
0
400
sample 8
0
400
sample 9
0
400
sample 10
0
400
sample 11
0
400
sample 12
0
400
sample 13
0
400
sample 14
0
400
sample 15
0
400
sample 16
0
400
sample 17
0
400
sample 18
0
400
sample 19
0
400
sample 20
0
400
sample 21
0
400
sample 22
0
400
sample 23
0
400
sample 24
0
400
sample 25
0
400
sample 26
0
400
sample 27
0
400
sample 28
0
400
sample 29
0
400
sample 30
0
400
sample 31
0
400
sample 32
0
400
sample 33
0
400
sample 34
0
400

Figure 4b

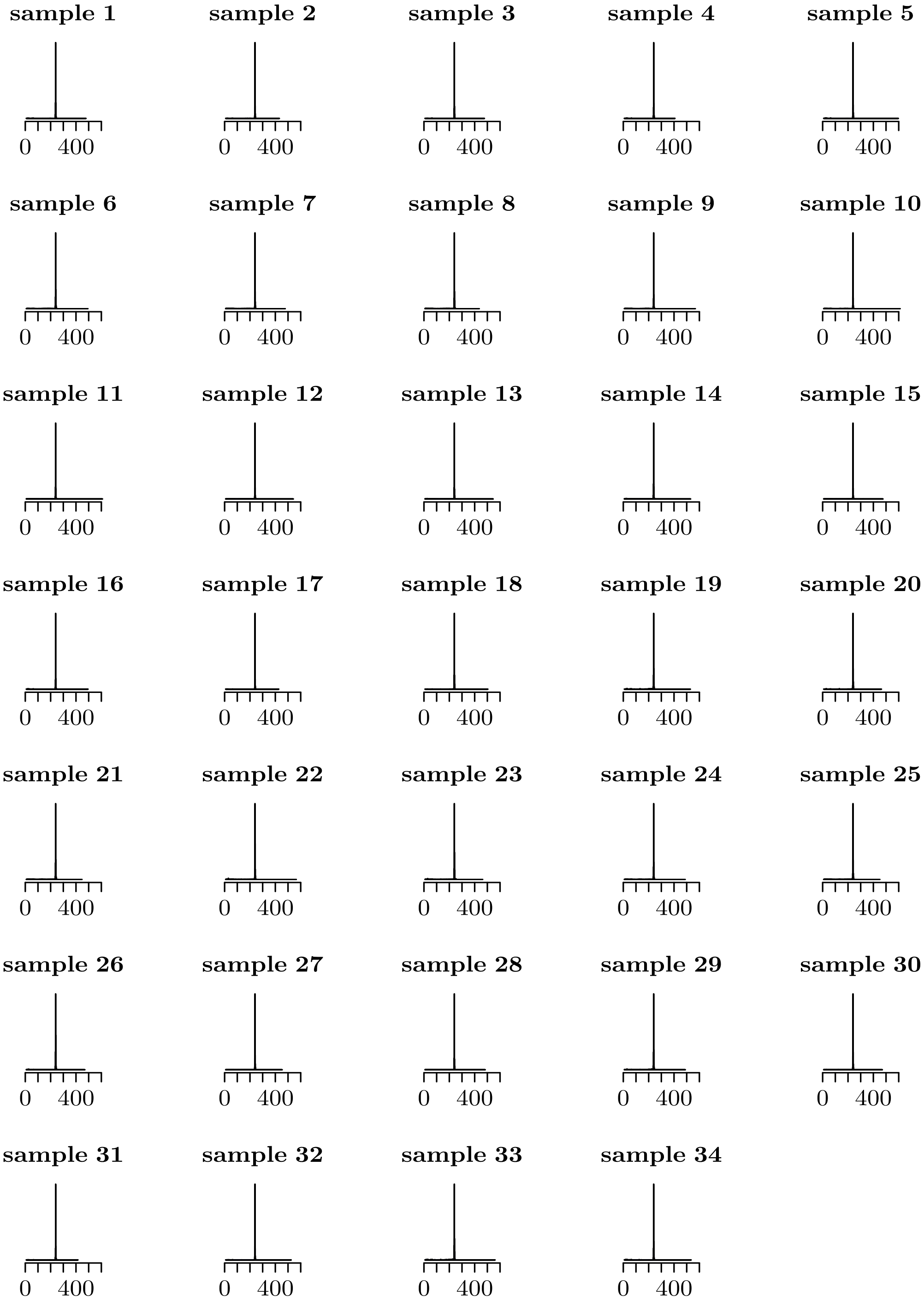


Figure 4c

Figure 4: Read length distribution for 16S data, chips 1 (4a), 2 (4b), 3 (4c).

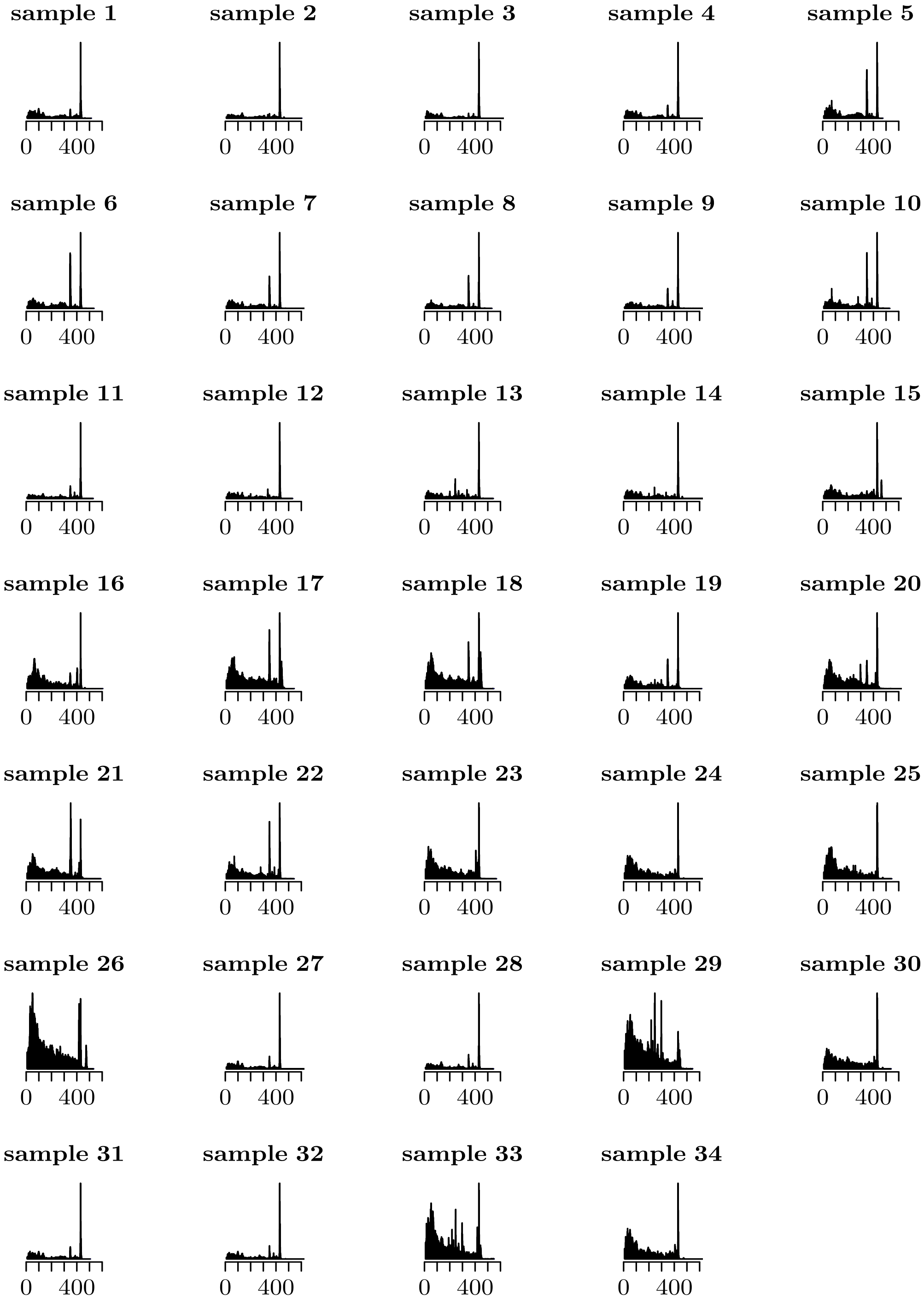

sample 1
0 400
sample 2
0 400
sample 3
0 400
sample 4
0 400
sample 5
0 400
sample 6
0 400
sample 7
0 400
sample 8
0 400
sample 9
0 400
sample 10
0 400
sample 11
0 400
sample 12
0 400
sample 13
0 400
sample 14
0 400
sample 15
0 400
sample 16
0 400
sample 17
0 400
sample 18
0 400
sample 19
0 400
sample 20
0 400
sample 21
0 400
sample 22
0 400
sample 23
0 400
sample 24
0 400
sample 25
0 400
sample 26
0 400
sample 27
0 400
sample 28
0 400
sample 29
0 400
sample 30
0 400
sample 31
0 400
sample 32
0 400
sample 33
0 400
sample 34
0 400

Figure 5a

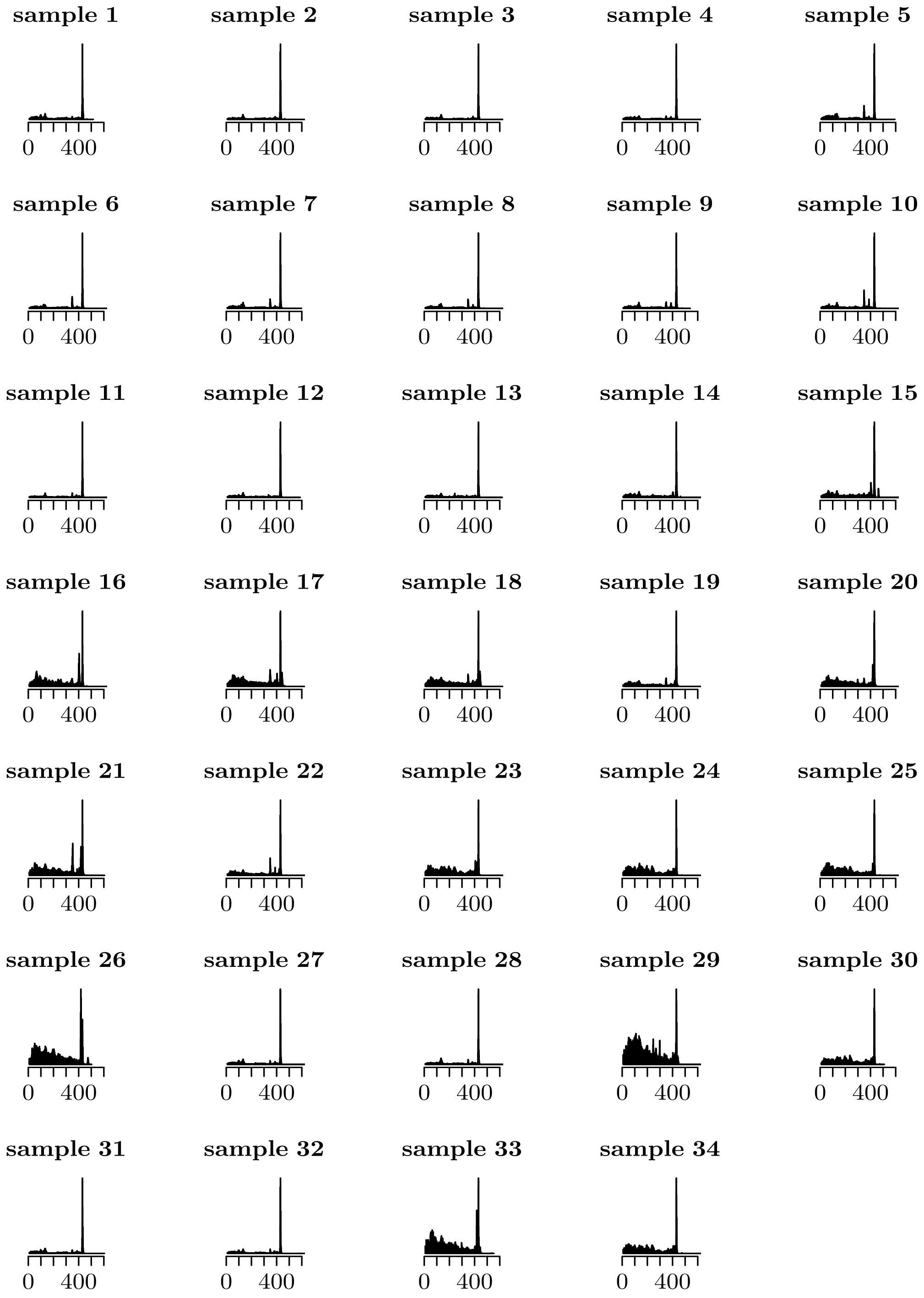

sample 1
0 400
sample 2
0 400
sample 3
0 400
sample 4
0 400
sample 5
0 400
sample 6
0 400
sample 7
0 400
sample 8
0 400
sample 9
0 400
sample 10
0 400
sample 11
0 400
sample 12
0 400
sample 13
0 400
sample 14
0 400
sample 15
0 400
sample 16
0 400
sample 17
0 400
sample 18
0 400
sample 19
0 400
sample 20
0 400
sample 21
0 400
sample 22
0 400
sample 23
0 400
sample 24
0 400
sample 25
0 400
sample 26
0 400
sample 27
0 400
sample 28
0 400
sample 29
0 400
sample 30
0 400
sample 31
0 400
sample 32
0 400
sample 33
0 400
sample 34
0 400

Figure 5b

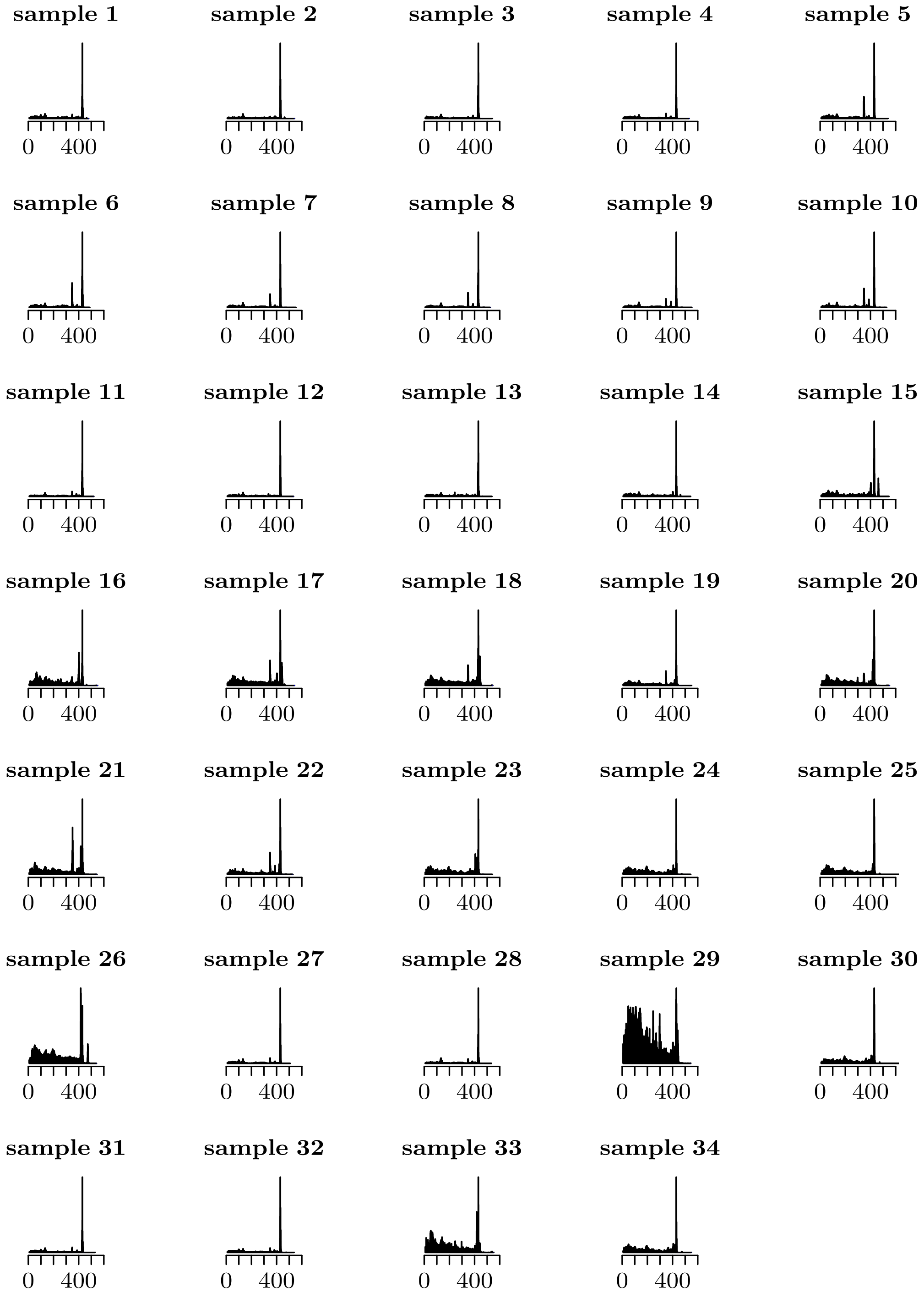

sample 1
0
400
sample 2
0
400
sample 3
0
400
sample 4
0
400
sample 5
0
400
sample 6
0
400
sample 7
0
400
sample 8
0
400
sample 9
0
400
sample 10
0
400
sample 11
0
400
sample 12
0
400
sample 13
0
400
sample 14
0
400
sample 15
0
400
sample 16
0
400
sample 17
0
400
sample 18
0
400
sample 19
0
400
sample 20
0
400
sample 21
0
400
sample 22
0
400
sample 23
0
400
sample 24
0
400
sample 25
0
400
sample 26
0
400
sample 27
0
400
sample 28
0
400
sample 29
0
400
sample 30
0
400
sample 31
0
400
sample 32
0
400
sample 33
0
400
sample 34
0
400

Figure 5c

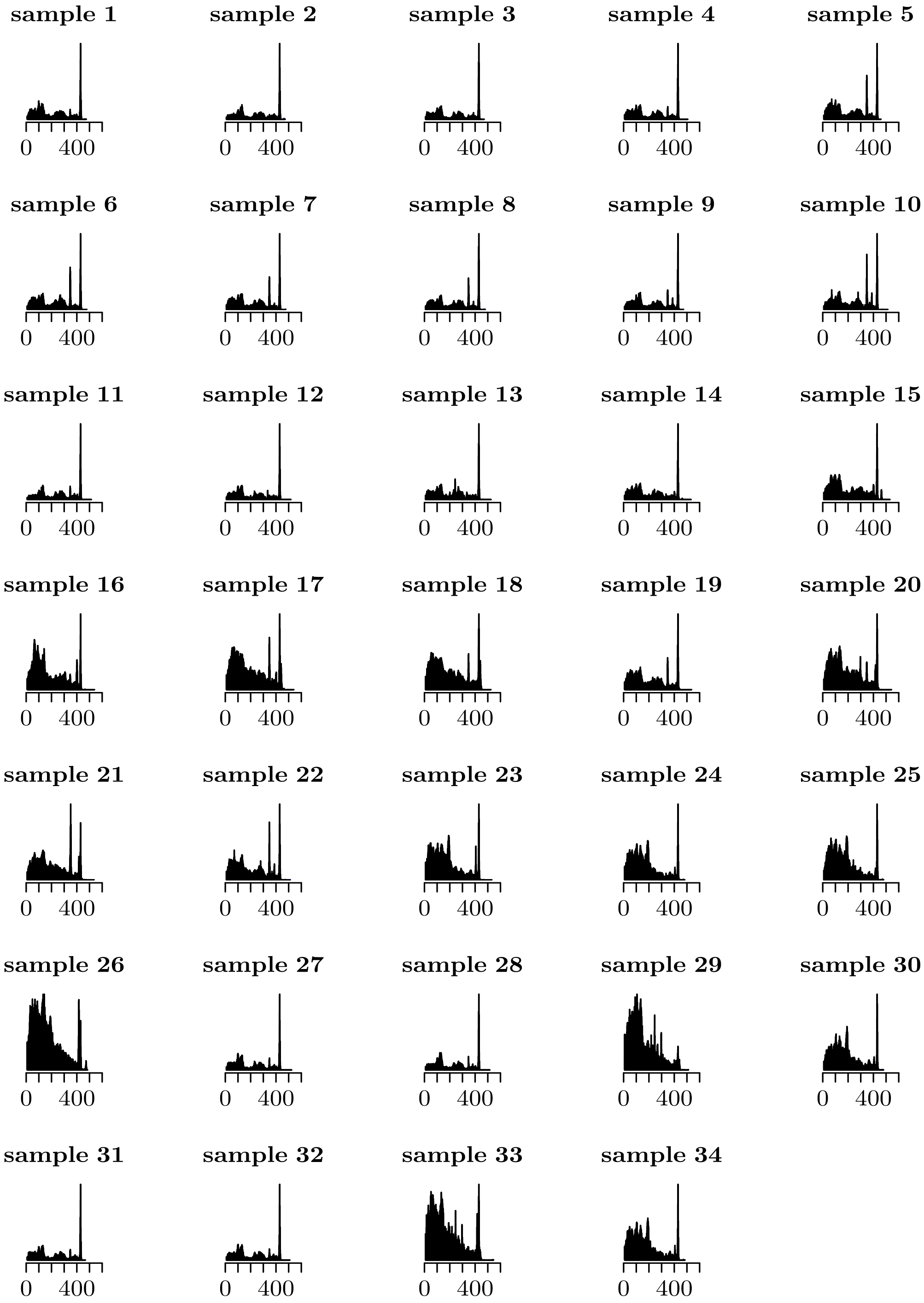
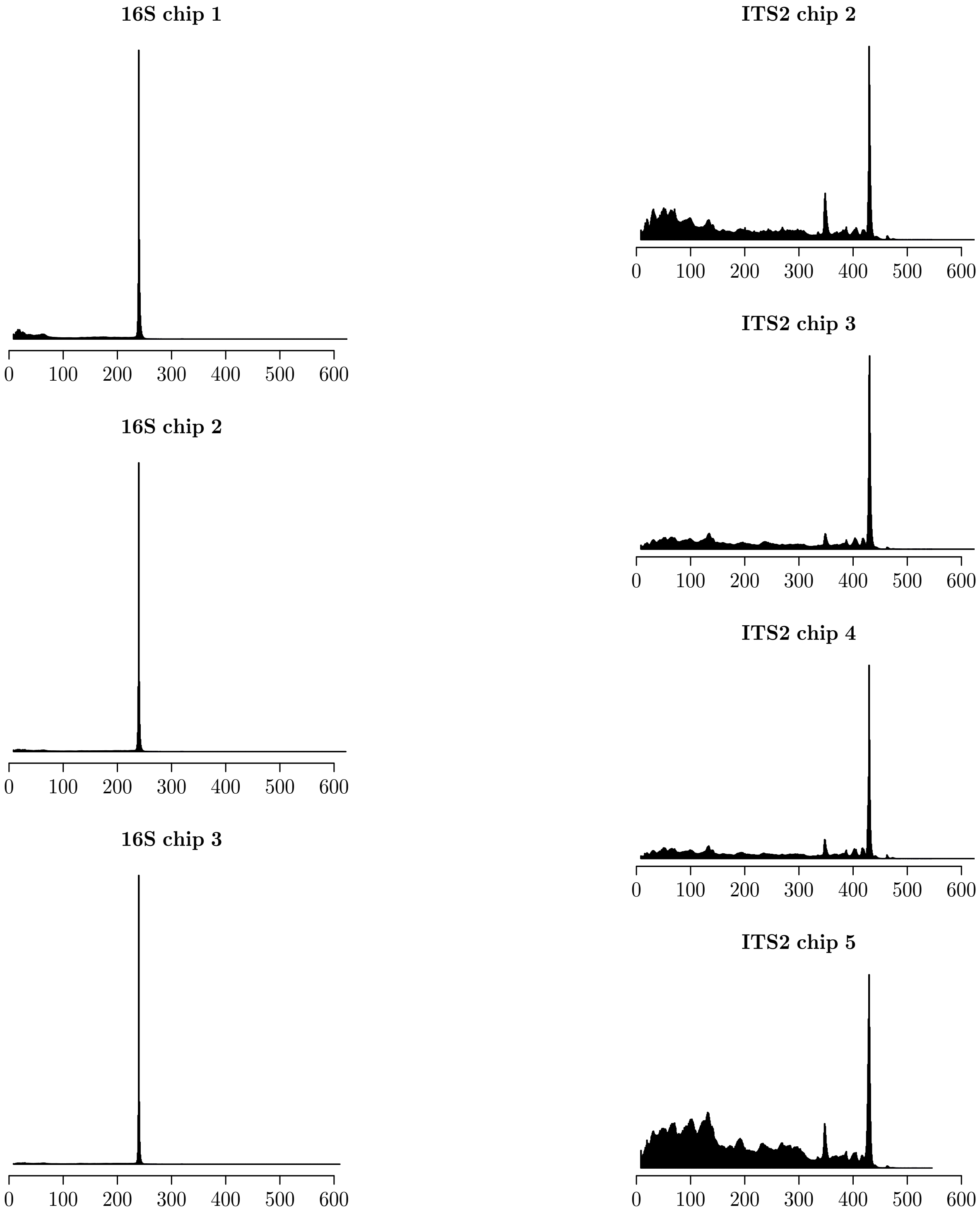


Figure 5d

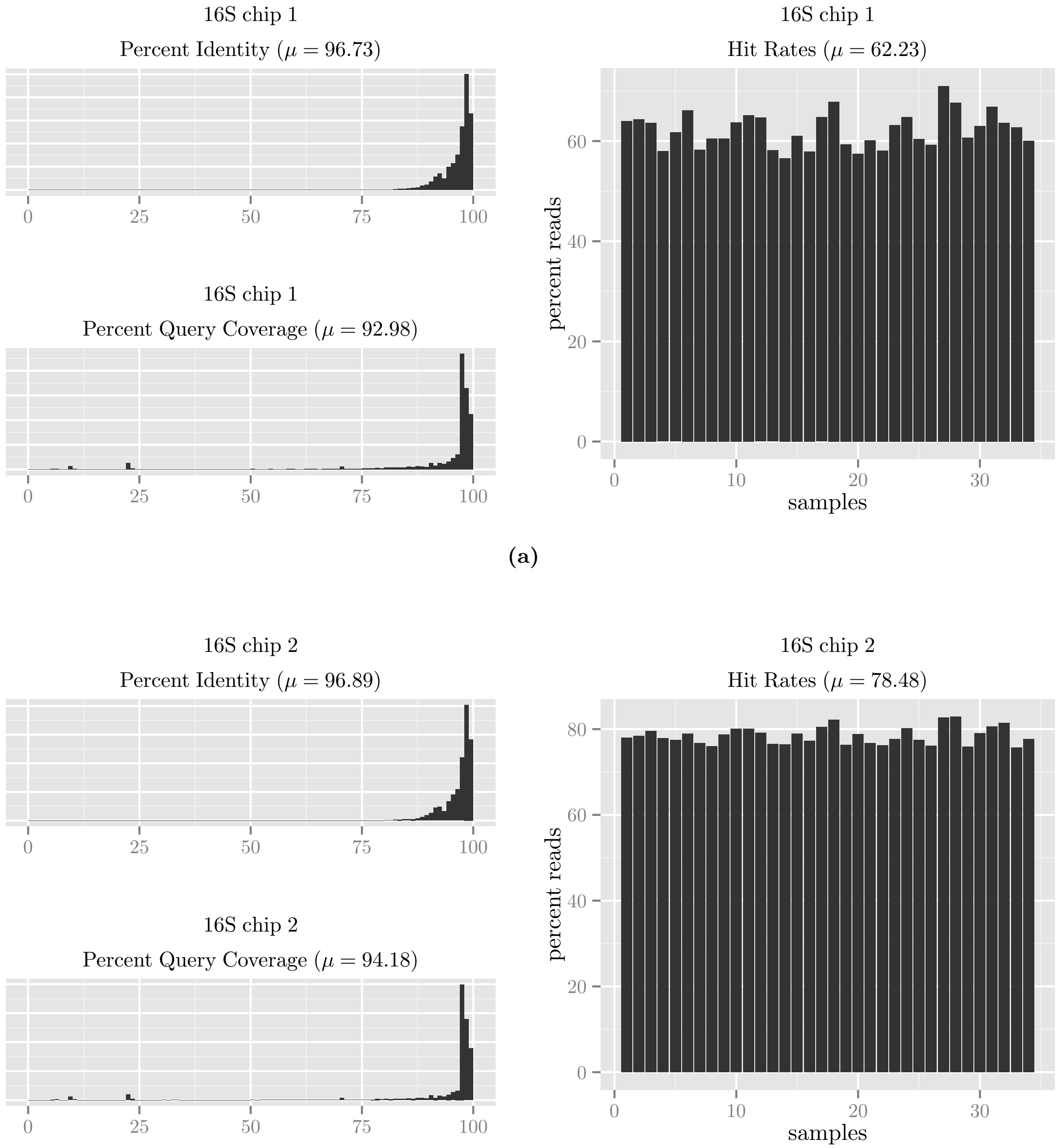
Figure 5: Read length distribution for ITS2 data, chips 2 (5a), 3 (5b), 4 (5c), 5 (5d).



(a)

(b)

Figure 6: Read length distributions for all 16S (6a) and ITS2 (6b) data.



16S chip 1

Percent Identity ($\mu = 96.73$)

16S chip 1

Percent Query Coverage ($\mu = 92.98$)

16S chip 1

Hit Rates ($\mu = 62.23$)

(a)

16S chip 2

Percent Identity ($\mu = 96.89$)

16S chip 2

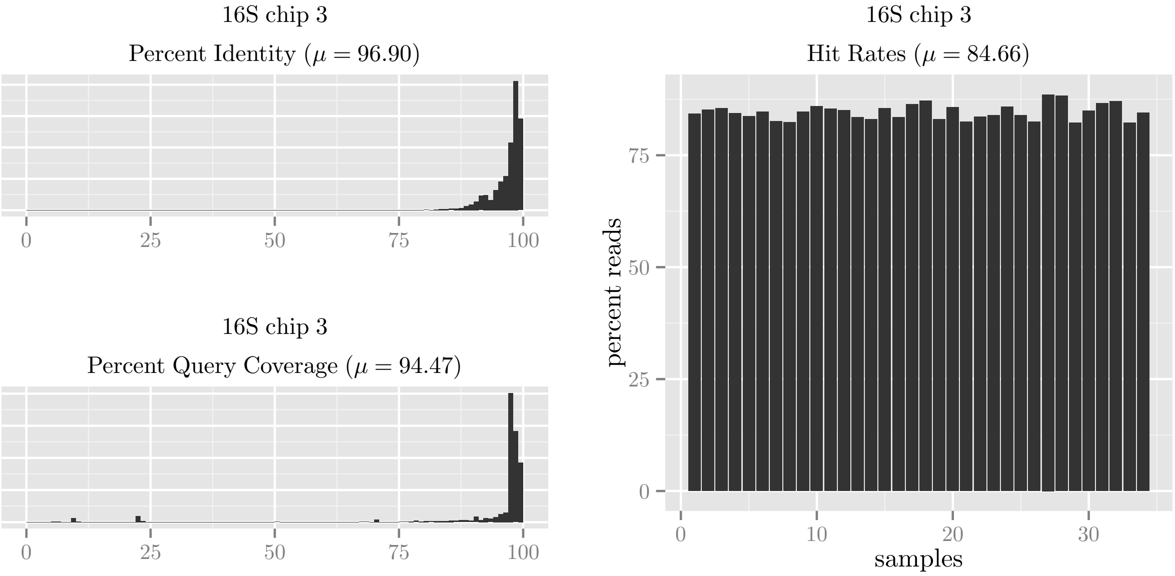
Percent Query Coverage ($\mu = 94.18$)

16S chip 2

Hit Rates ($\mu = 78.48$)

(b)

Figures 7a and 7b



16S chip 3

Percent Identity ($\mu = 96.90$)

16S chip 3

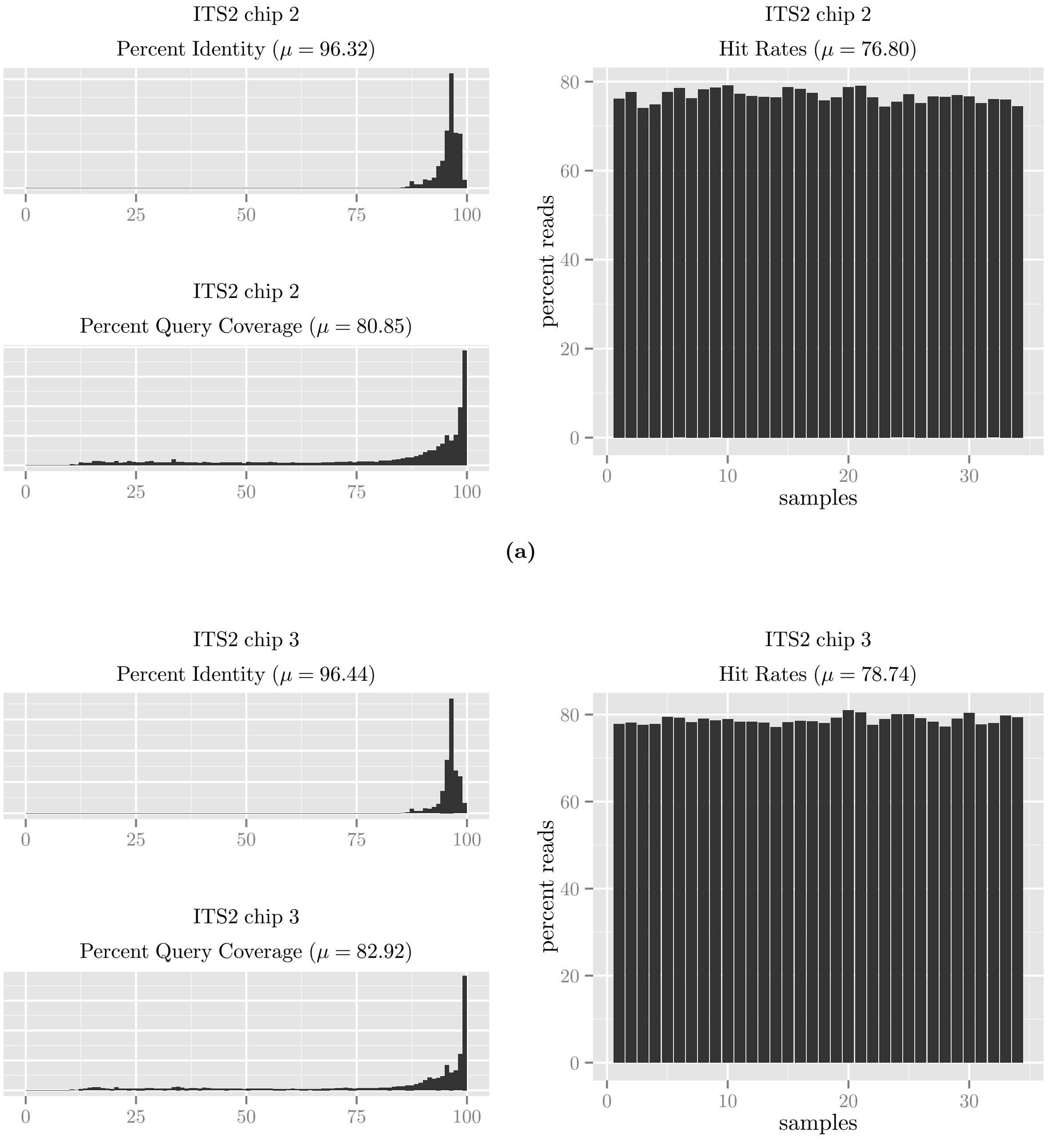
Percent Query Coverage ($\mu = 94.47$)

16S chip 3

Hit Rates ($\mu = 84.66$)

Figure 7c

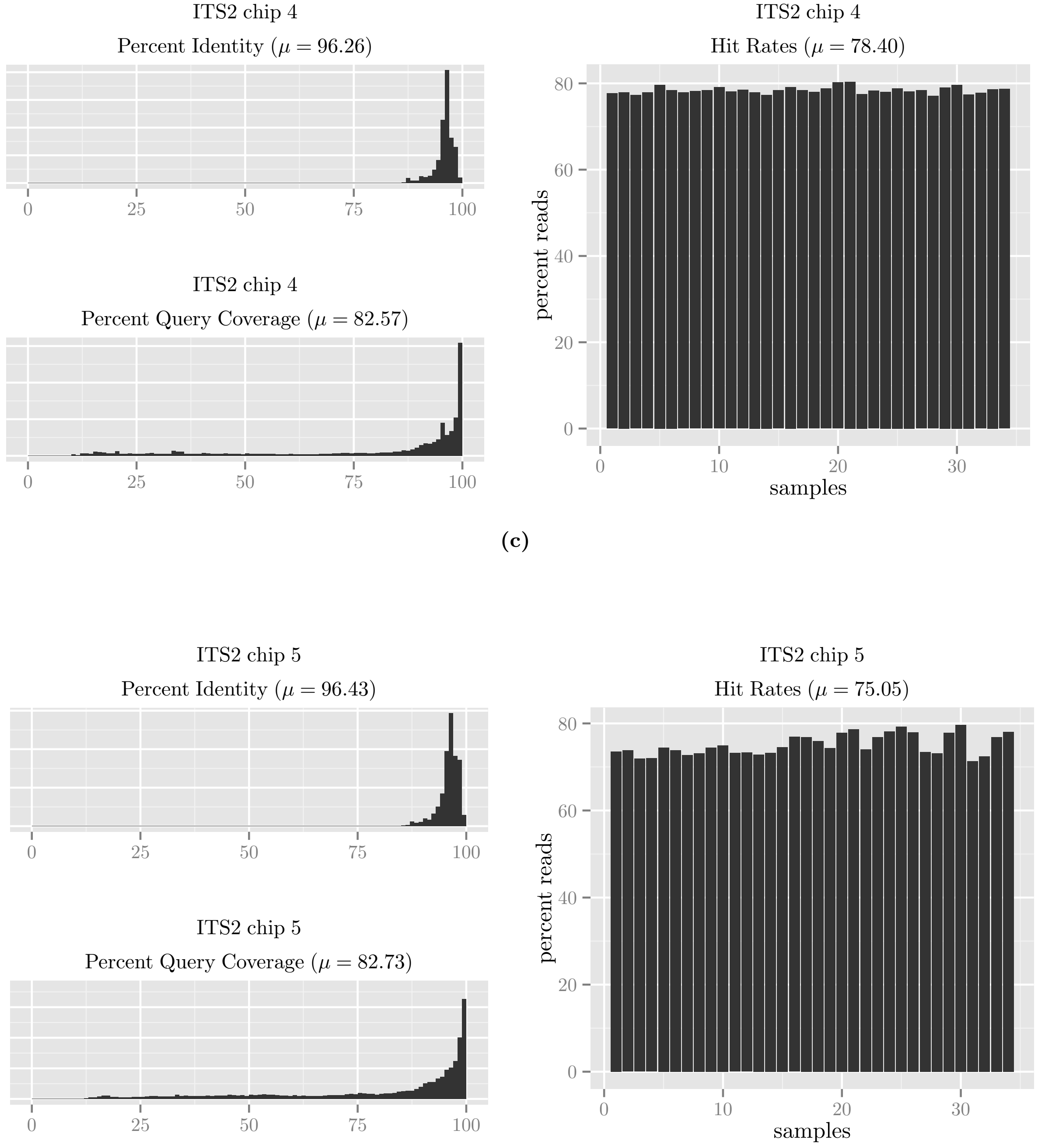
Figure 7: Figures 7a, 7b, 7c shows the percent identities (%ID) and query coverages (%COV) of mapping sequences for chips 1, 2, 3; together with the percentages of sequences that are accepted as hit, after applying the 80% and 90% %COV and %ID cutoffs for all 34 samples.



(a)

(b)

Figures 8a and 8b

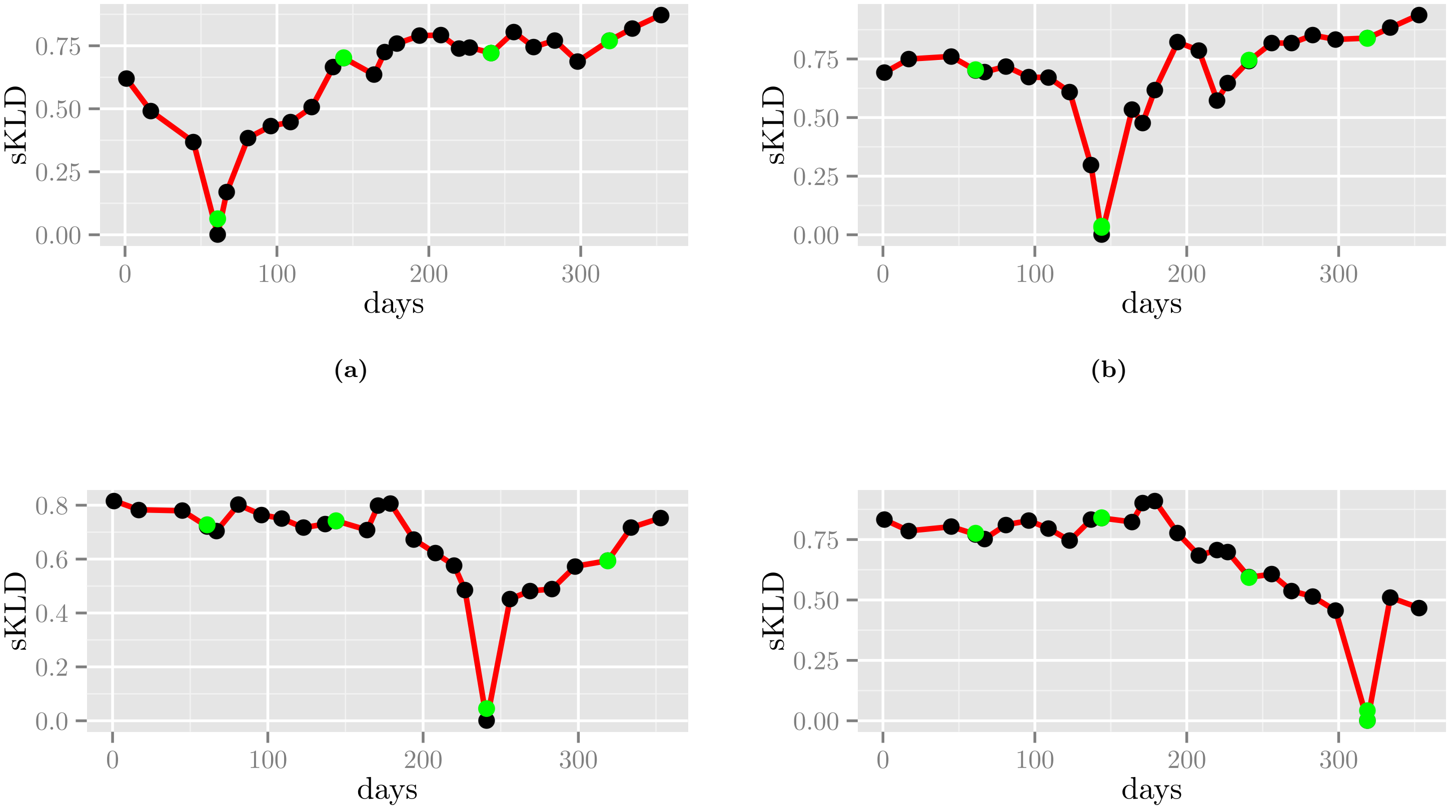


(c)

(d)

Figures 8c and 8d

Figure 8: Figures 8a, 8b, 8c, 8d shows the percent identities (%ID) and query coverages (%COV) of mapping sequences for chips 2, 3, 4, 5; together with the percentages of sequences that are accepted as hit, after applying the 80% and 90% %COV and %ID cutoffs for all 34 samples.



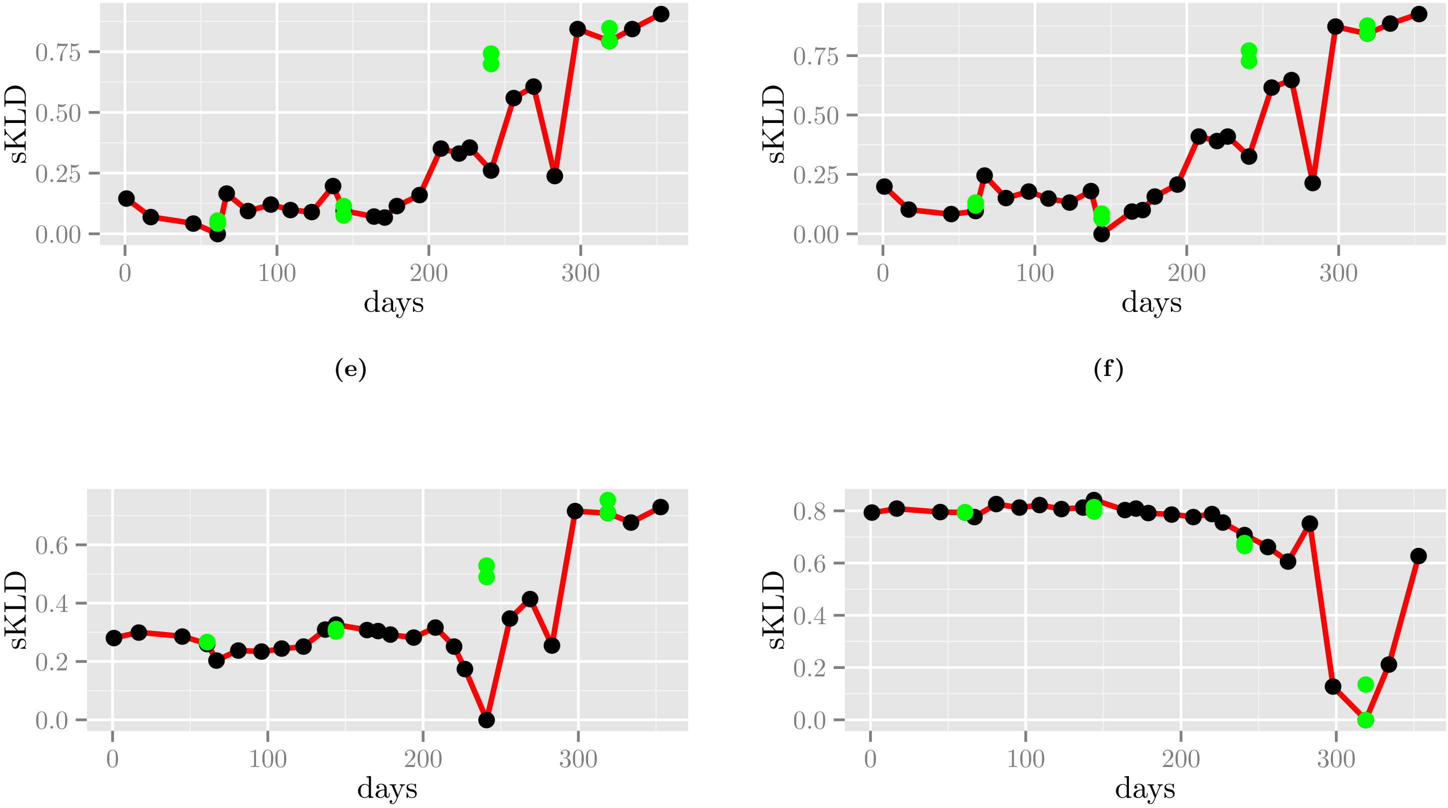
(a)

(b)

(c)

(d)

Figures 9a, 9b, 9c, and 9d



(e)

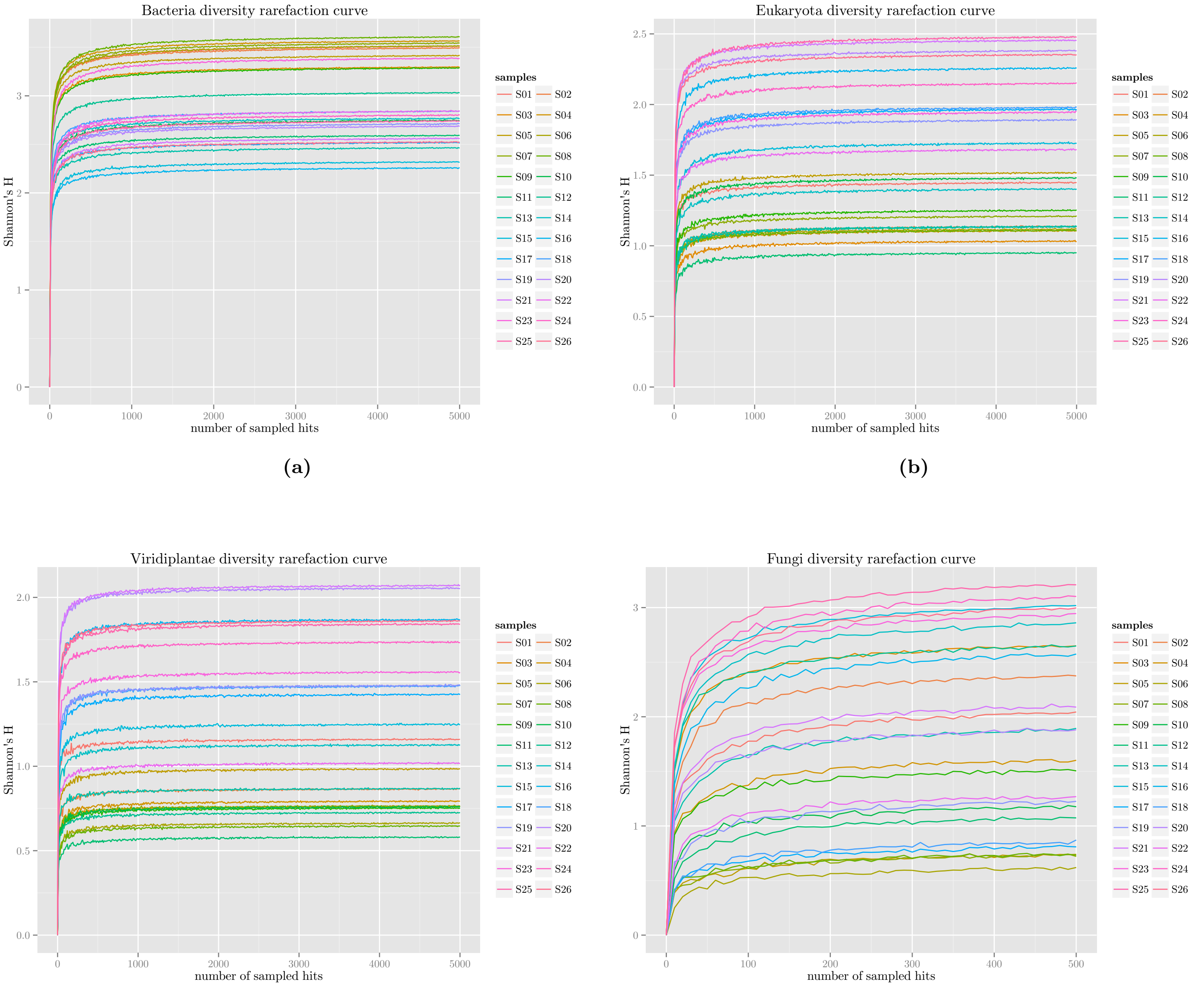
(f)

(g)

(h)

Figures 9e, 9f, 9g, and 9h

Figure 9: Divergences across selected samples: 9a, 9b, 9c, and 9d shows the distances between sample 4, 11, 19, 24, and all other samples, respectively for 16S data, whereas 9e, 9f, 9g, and 9h shows it for ITS2 data. Grey points correspond to original samples, while green points represent the technical replicates of the samples sharing their x-axis value. The zero KL distance (y-axis) on each plot indicates which sample all other samples are compared against. Good reproducibility is achieved when the green points superimposed over the fixed samples (4, 11, 19, 24) also have zero KLD values.



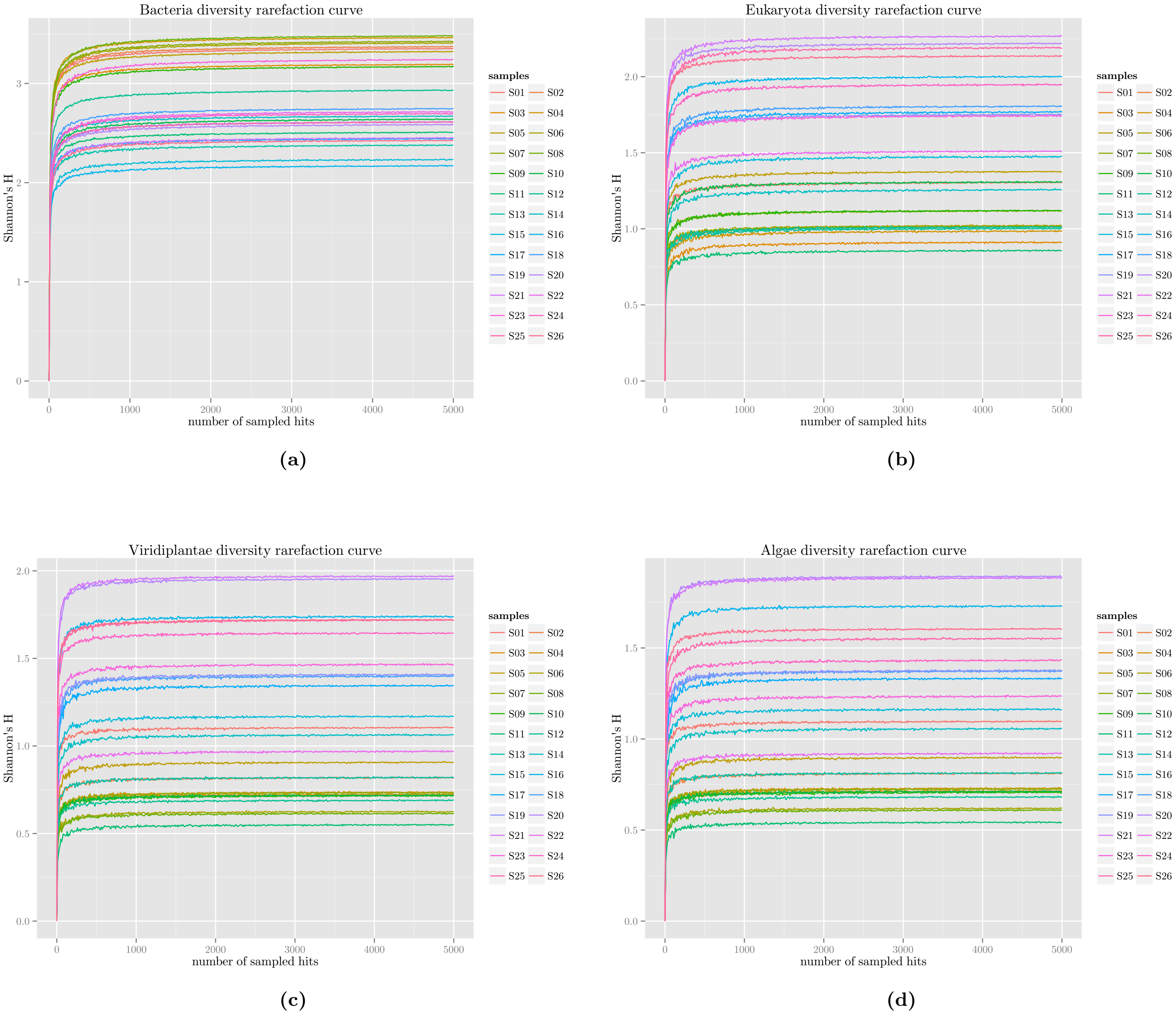
(a)

(b)

(c)

(d)

Figure 10: Rarefaction Curves: Depicts the converging diversity (Shannon H) rarefaction curves for Bacteria, Eukaryota, Viridiplantae, algae, and Fungi, over all 16S and ITS2 reference sequences, averaged over 100 interations.



(a)

(b)

(c)

(d)

Figures 11a, 11b, 11c and 11d

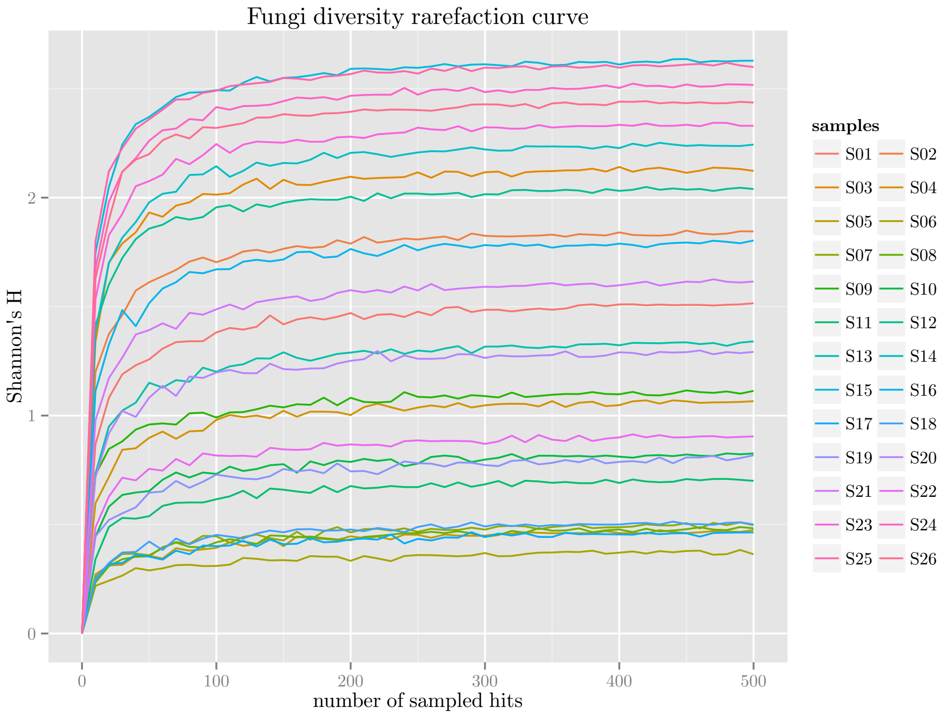


Figure 11e

Figure 11: Rarefaction Curves: Depicts the converging diversity (Shannon H) rarefaction curves for Bacteria, Eukaryota, Viridiplantae, algae, and Fungi, over the top 2000 and 200 16S and ITS2 reference sequences, averaged over 100 interations.

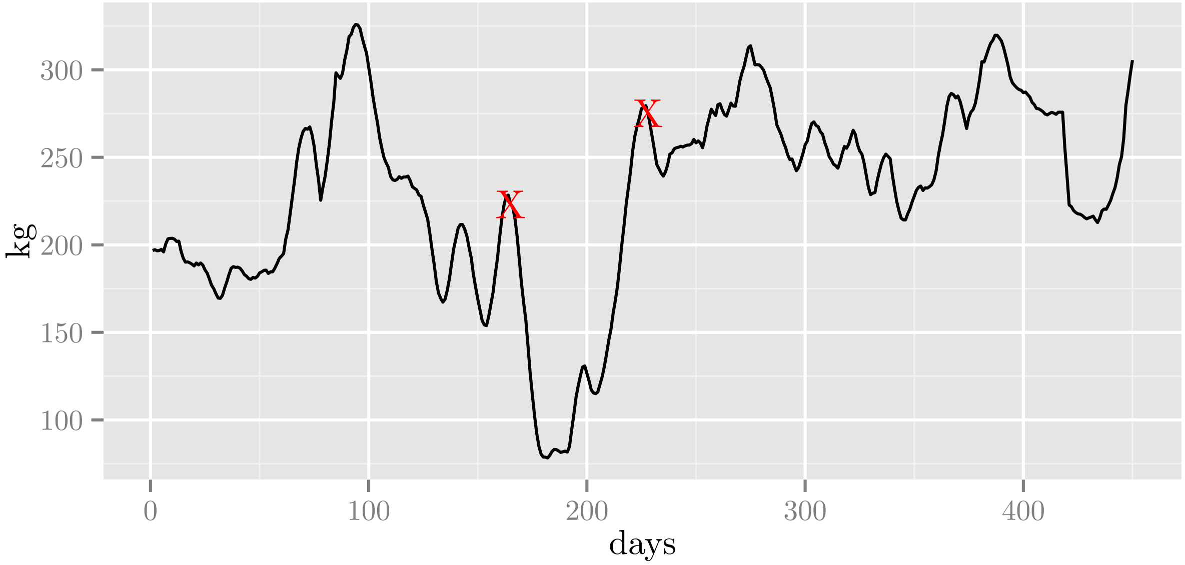
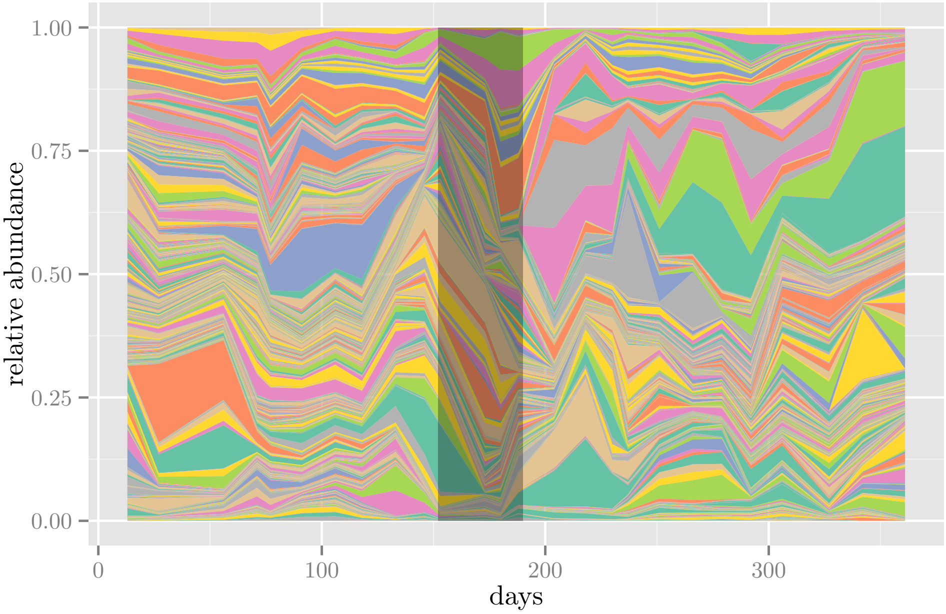
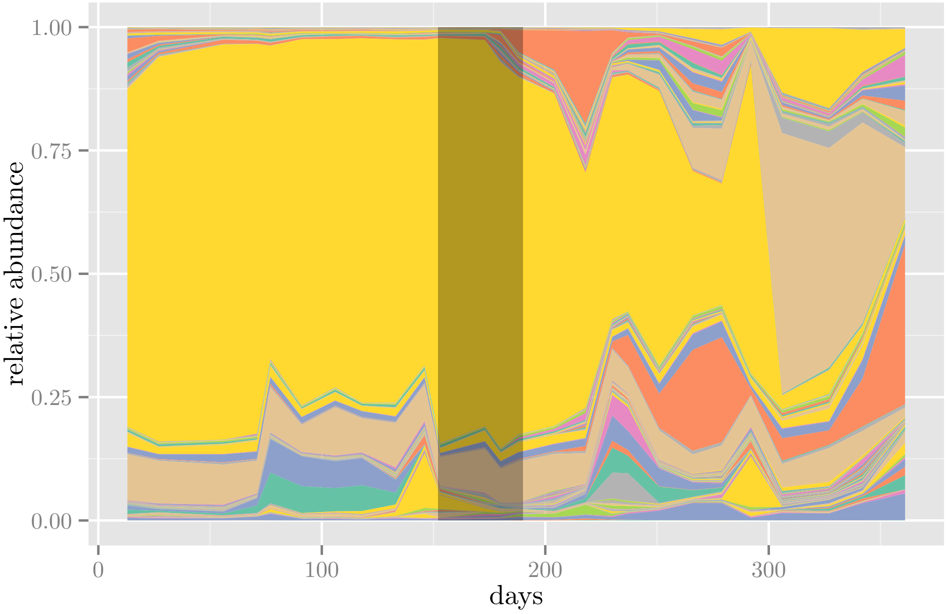


Figure 12: Dry weight (kg): Algal dry weight in kg, with peaks on days 165, and 228 marked.

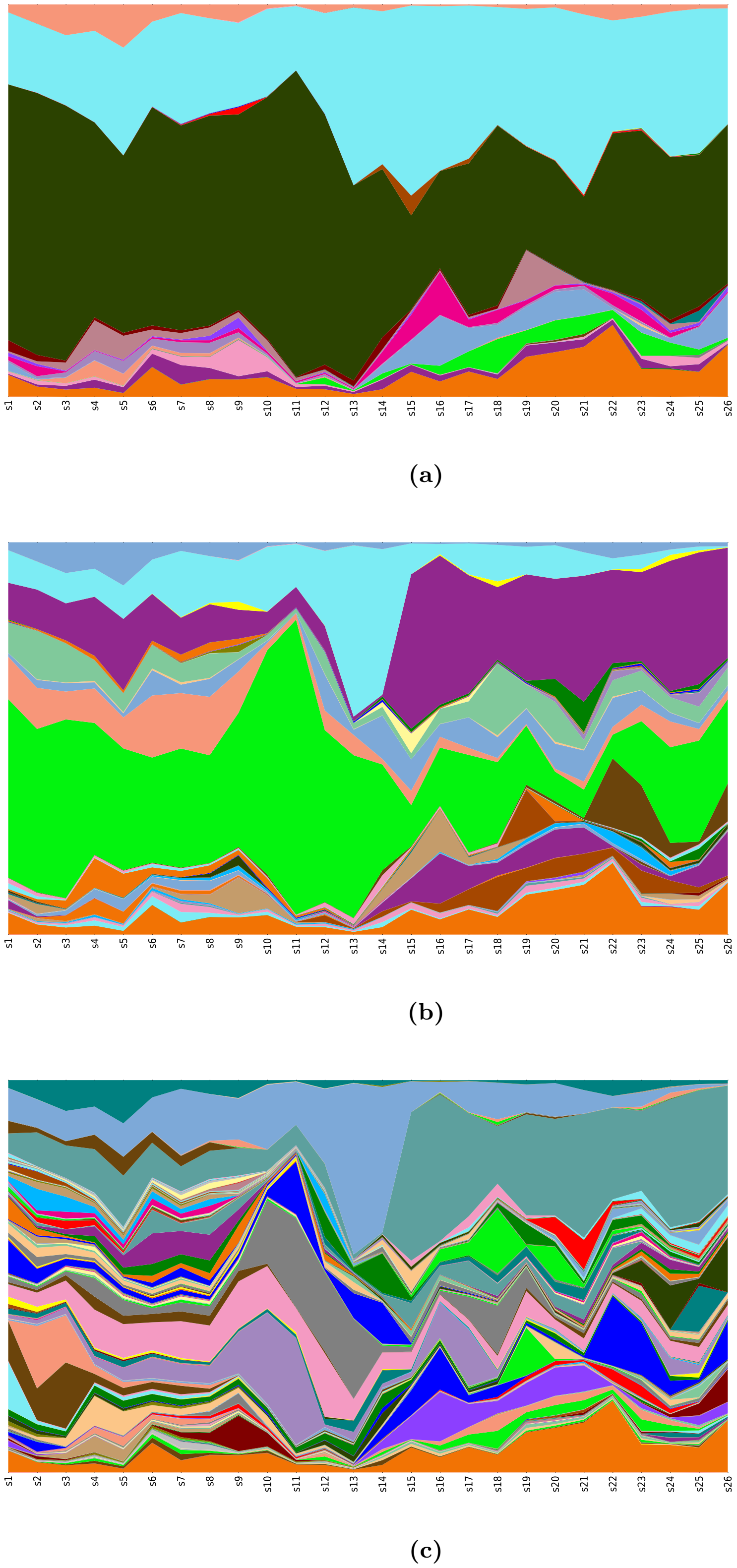


(a) Top 1000 sequences hit in GreenGenes.



(b) Top 200 sequences hit in constructed ITS2 database from NCBI.

Figure 13: Finest granularity (sequence level) area plots: Top hit reference sequences in 16S, using two different databases, and ITS2 data, respectively.

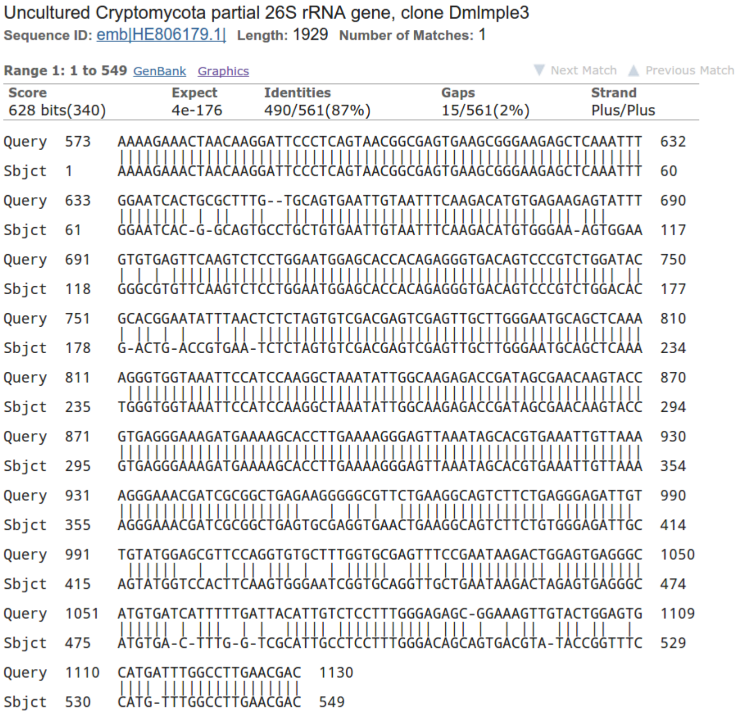


(a)

(b)

(c)

Figure 14: Bralizian Microbiome Pipeline area plots at phylum (14a), class (14b), and genus (14c) levels for 16S data. Taxa not shown.



Uncultured Cryptomycota partial 26S rRNA gene, clone DmImple3

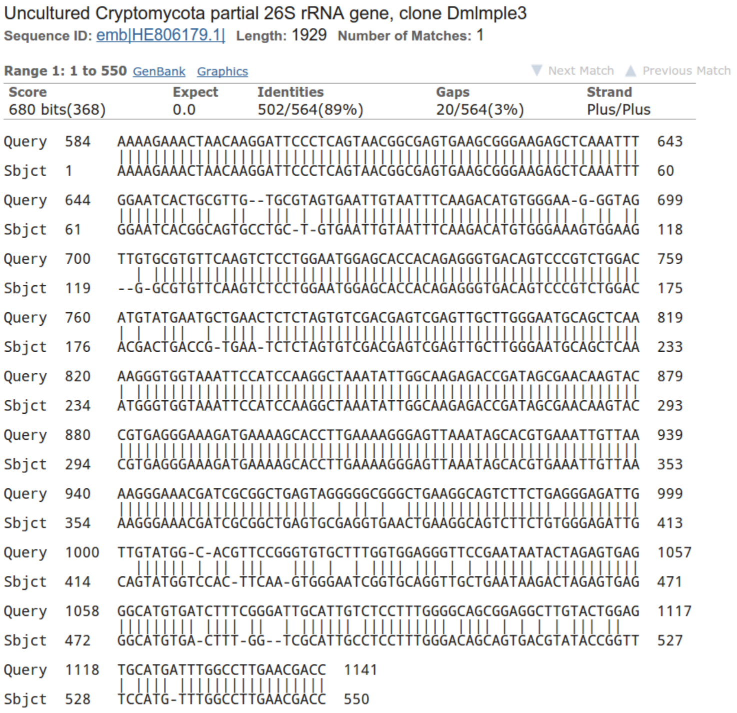
Sequence ID: emb|HE806179.1| Length: 1929 Number of Matches: 1

Range 1: 1 to 549 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
628 bits(340)	4e-176	490/561(87%)	15/561(2%)	Plus/Plus

```
Query  573   AAAAGAAACTAACAAGGATTCCCTCAGTAACGGCGAGTGAAGCGGGAAGAGCTCAAATTT  632
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1     AAAAGAAACTAACAAGGATTCCCTCAGTAACGGCGAGTGAAGCGGGAAGAGCTCAAATTT  60

Query  633   GGAATCACTGCGCTTTG--TGCAGTGAATTGTAATTTCAAGACATGTGAGAAGAGTATTT  690
             |||||||| | ||  ||  ||| ||||||||||||||||||||||||| ||| |||
Sbjct  61    GGAATCAC-G-GCAGTGCCTGCTGTGAATTGTAATTTCAAGACATGTGGGAA-AGTGGAA  117

Query  691   GTGTGAGTTCAAGTCTCCTGGAATGGAGCACCACAGAGGGTGACAGTCCCGTCTGGATAC  750
             | | | ||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Sbjct  118   GGGCGTGTTCAAGTCTCCTGGAATGGAGCACCACAGAGGGTGACAGTCCCGTCTGGACAC  177

Query  751   GCACGGAATATTTAACTCTCTAGTGTCGACGAGTCGAGTTGCTTGGGAATGCAGCTCAAA  810
             | || | |   | || ||||||||||||||||||||||||||||||||||||||||||||
Sbjct  178   G-ACTG-ACCGTGAA-TCTCTAGTGTCGACGAGTCGAGTTGCTTGGGAATGCAGCTCAAA  234

Query  811   AGGGTGGTAAATTCCATCCAAGGCTAAATATTGGCAAGAGACCGATAGCGAACAAGTACC  870
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  235   TGGGTGGTAAATTCCATCCAAGGCTAAATATTGGCAAGAGACCGATAGCGAACAAGTACC  294

Query  871   GTGAGGGAAAGATGAAAAGCACCTTGAAAAGGGAGTTAAATAGCACGTGAAATTGTTAAA  930
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  295   GTGAGGGAAAGATGAAAAGCACCTTGAAAAGGGAGTTAAATAGCACGTGAAATTGTTAAA  354

Query  931   AGGGAAACGATCGCGGCTGAGAAGGGGGCGTTCTGAAGGCAGTCTTCTGAGGGAGATTGT  990
             |||||||||||||||||||||   | || |  ||||||||||||||||| |||||||||
Sbjct  355   AGGGAAACGATCGCGGCTGAGTGCGAGGTGAACTGAAGGCAGTCTTCTGTGGGAGATTGC  414

Query  991   TGTATGGAGCGTTCCAGGTGTGCTTTGGTGCGAGTTTCCGAATAAGACTGGAGTGAGGGC  1050
              ||||||  |  | || ||| |  | |||||  ||| | |||||||||| ||||||||||
Sbjct  415   AGTATGGTCCACTTCAAGTGGGAATCGGTGCAGGTTGCTGAATAAGACTAGAGTGAGGGC  474

Query  1051  ATGTGATCATTTTTGATTACATTGTCTCCTTTGGGAGAGC-GGAAAGTTGTACTGGAGTG  1109
             |||||| | |||  | |  ||||| ||||||||||| ||| |  | ||  ||| ||  |
Sbjct  475   ATGTGA-C-TTTG-G-TCGCATTGCCTCCTTTGGGACAGCAGTGACGTA-TACCGGTTTC  529

Query  1110  CATGATTTGGCCTTGAACGAC  1130
             |||| ||||||||||||||||
Sbjct  530   CATG-TTTGGCCTTGAACGAC  549
```

(a) Alignment of GI: 532165669



Uncultured Cryptomycota partial 26S rRNA gene, clone DmImple3

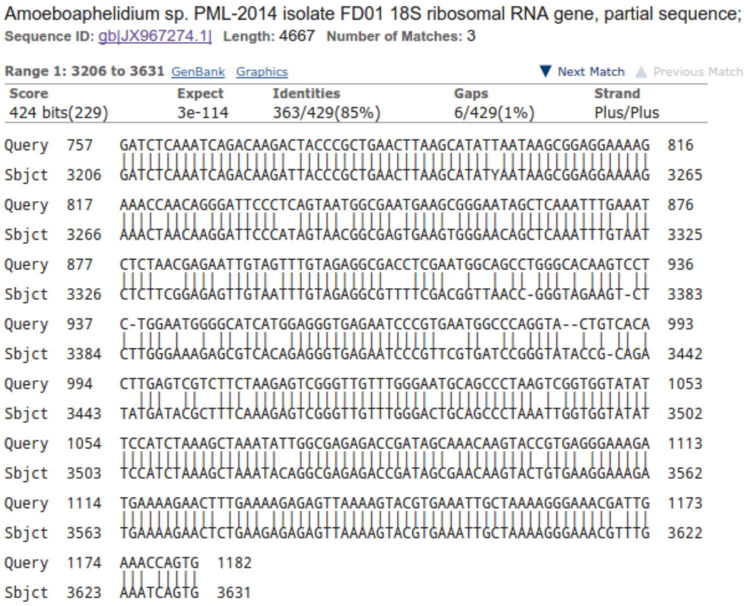
Sequence ID: emb|HE806179.1| Length: 1929 Number of Matches: 1

Range 1: 1 to 550 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
680 bits(368)	0.0	502/564(89%)	20/564(3%)	Plus/Plus

```
Query  584   AAAAGAAACTAACAAGGATTCCCTCAGTAACGGCGAGTGAAGCGGGAAGAGCTCAAATTT  643
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1     AAAAGAAACTAACAAGGATTCCCTCAGTAACGGCGAGTGAAGCGGGAAGAGCTCAAATTT  60

Query  644   GGAATCACTGCGTTG--TGCGTAGTGAATTGTAATTTCAAGACATGTGGGAA-G-GGTAG  699
             |||||||| ||  ||  ||| | ||||||||||||||||||||||||||||| | || ||
Sbjct  61    GGAATCACGGCAGTGCCTGC-T-GTGAATTGTAATTTCAAGACATGTGGGAAAGTGGAAG  118

Query  700   TTGTGCGTGTTCAAGTCTCCTGGAATGGAGCACCACAGAGGGTGACAGTCCCGTCTGGAC  759
               | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  119   --G-GCGTGTTCAAGTCTCCTGGAATGGAGCACCACAGAGGGTGACAGTCCCGTCTGGAC  175

Query  760   ATGTATGAATGCTGAACTCTCTAGTGTCGACGAGTCGAGTTGCTTGGGAATGCAGCTCAA  819
             | |  |||  | |||| |||||||||||||||||||||||||||||||||||||||||||
Sbjct  176   ACGACTGACCG-TGAA-TCTCTAGTGTCGACGAGTCGAGTTGCTTGGGAATGCAGCTCAA  233

Query  820   AAGGGTGGTAAATTCCATCCAAGGCTAAATATTGGCAAGAGACCGATAGCGAACAAGTAC  879
             | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  234   ATGGGTGGTAAATTCCATCCAAGGCTAAATATTGGCAAGAGACCGATAGCGAACAAGTAC  293

Query  880   CGTGAGGGAAAGATGAAAAGCACCTTGAAAAGGGAGTTAAATAGCACGTGAAATTGTTAA  939
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  294   CGTGAGGGAAAGATGAAAAGCACCTTGAAAAGGGAGTTAAATAGCACGTGAAATTGTTAA  353

Query  940   AAGGGAAACGATCGCGGCTGAGTAGGGGGCGGGCTGAAGGCAGTCTTCTGAGGGAGATTG  999
             |||||||||||||||||||||||  | || |  ||||||||||||||||| |||||||||
Sbjct  354   AAGGGAAACGATCGCGGCTGAGTGCGAGGTGAACTGAAGGCAGTCTTCTGTGGGAGATTG  413

Query  1000  TTGTATGG-C-ACGTTCCGGGTGTGCTTTGGTGGAGGGTTCCGAATAATACTAGAGTGAG  1057
               |||||| | || |||   ||| |  | |||| ||| | | |||||| |||||||||||
Sbjct  414   CAGTATGGTCCAC-TTCAA-GTGGGAATCGGTGCAGGTTGCTGAATAAGACTAGAGTGAG  471

Query  1058  GGCATGTGATCTTTCGGGATTGCATTGTCTCCTTTGGGGCAGCGGAGGCTTGTACTGGAG  1117
             ||||||||| |||| ||  | |||||| |||||||||| |||| | | | | ||| ||
Sbjct  472   GGCATGTGA-CTTT-GG--TCGCATTGCCTCCTTTGGGACAGCAGTGACGTATACCGGTT  527

Query  1118  TGCATGATTTGGCCTTGAACGACC  1141
             | |||| |||||||||||||||||
Sbjct  528   TCCATG-TTTGGCCTTGAACGACC  550
```

(b) Alignment of GI: 532165968



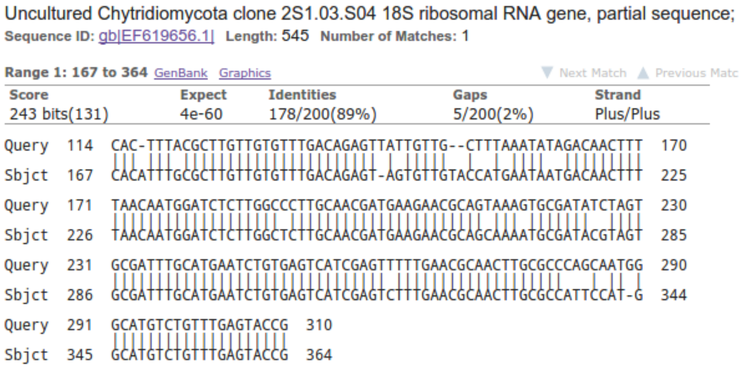
Amoeboaphelidium sp. PML-2014 isolate FD01 18S ribosomal RNA gene, partial sequence;

Sequence ID: gb|JX967274.1| Length: 4667 Number of Matches: 3

Range 1: 3206 to 3631 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
424 bits(229)	3e-114	363/429(85%)	6/429(1%)	Plus/Plus

```
Query  757   GATCTCAAATCAGACAAGACTACCCGCTGAACTTAAGCATATTAATAAGCGGAGGAAAAG  816
             ||||||||||||||||||| |||||||||||||||||||||||| |||||||||||||||||
Sbjct  3206  GATCTCAAATCAGACAAGATTACCCGCTGAACTTAAGCATATYAATAAGCGGAGGAAAAG  3265

Query  817   AAACCAACAGGGATTCCCTCAGTAATGGCGAATGAAGCGGGAATAGCTCAAATTTGAAAT  876
             |||| |||| ||||||||  ||||| ||||| ||||| ||||| |||||||||||| |||
Sbjct  3266  AAACTAACAAGGATTCCCATAGTAACGGCGAGTGAAGTGGGAACAGCTCAAATTTGTAAT  3325

Query  877   CTCTAACGAGAATTGTAGTTTGTAGAGGCGACCTCGAATGGCAGCCTGGGCACAAGTCCT  936
             ||||   |||| ||||| ||||||||||||   ||||  |  | || ||| | |||| ||
Sbjct  3326  CTCTTCGGAGAGTTGTAATTTGTAGAGGCGTTTTCGACGGTTAACC-GGGTAGAAGT-CT  3383

Query  937   C-TGGAATGGGGCATCATGGAGGGTGAGAATCCCGTGAATGGCCCAGGTA--CTGTCACA  993
             | ||| |  | || |||  |||||||||||||||||   ||  || ||||  | | || |
Sbjct  3384  CTTGGGAAAGAGCGTCACAGAGGGTGAGAATCCCGTTCGTGATCCGGGTATACCG-CAGA  3442

Query  994   CTTGAGTCGTCTTCTAAGAGTCGGGTTGTTTGGGAATGCAGCCCTAAGTCGGTGGTATAT  1053
               |||  ||  ||| |||||||||||||||||||| ||||||||||| | ||||||||||
Sbjct  3443  TATGATACGCTTTCAAAGAGTCGGGTTGTTTGGGACTGCAGCCCTAAATTGGTGGTATAT  3502

Query  1054  TCCATCTAAAGCTAAATATTGGCGAGAGACCGATAGCAAACAAGTACCGTGAGGGAAAGA  1113
             ||||||||||||||||||  ||||||||||||||||| ||||||||| |||| |||||||
Sbjct  3503  TCCATCTAAAGCTAAATACAGGCGAGAGACCGATAGCGAACAAGTACTGTGAAGGAAAGA  3562

Query  1114  TGAAAAGAACTTTGAAAAGAGAGTTAAAAGTACGTGAAATTGCTAAAAGGGAAACGATTG  1173
             ||||||||||| |||| ||||||||||||||||||||||||||||||||||||||| |||
Sbjct  3563  TGAAAAGAACTCTGAAGAGAGAGTTAAAAGTACGTGAAATTGCTAAAAGGGAAACGTTTG  3622

Query  1174  AAACCAGTG  1182
             ||| |||||
Sbjct  3623  AAATCAGTG  3631
```



Uncultured Chytridiomycota clone 2S1.03.S04 18S ribosomal RNA gene, partial sequence;

Sequence ID: gb|EF619656.1| Length: 545 Number of Matches: 1

Range 1: 167 to 364 GenBank Graphics ▼ Next Match ▲ Previous Matc

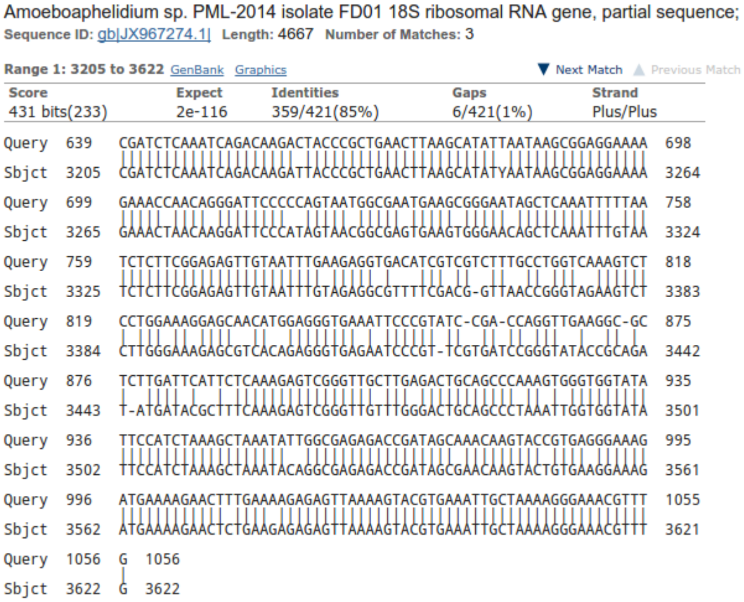
Score	Expect	Identities	Gaps	Strand
243 bits(131)	4e-60	178/200(89%)	5/200(2%)	Plus/Plus

```
Query  114   CAC-TTTACGCTTGTTGTGTTTGACAGAGTTATTGTTG--CTTTAAATATAGACAACTTT  170
             ||| ||| |||||||||||||||||||||| | |||||  |  | ||||  |||||||||
Sbjct  167   CACATTTGCGCTTGTTGTGTTTGACAGAGT-AGTGTTGTACCATGAATAATGACAACTTT  225

Query  171   TAACAATGGATCTCTTGGCCCTTGCAACGATGAAGAACGCAGTAAAGTGCGATATCTAGT  230
             ||||||||||||||||||| |||||||||||||||||||||| ||| |||||||  ||||
Sbjct  226   TAACAATGGATCTCTTGGCTCTTGCAACGATGAAGAACGCAGCAAAATGCGATACGTAGT  285

Query  231   GCGATTTGCATGAATCTGTGAGTCATCGAGTTTTTGAACGCAACTTGCGCCCAGCAATGG  290
             ||||||||||||||||||||||||||||||| |||||||||||||||||||   | || |
Sbjct  286   GCGATTTGCATGAATCTGTGAGTCATCGAGTCTTTGAACGCAACTTGCGCCATTCCAT-G  344

Query  291   GCATGTCTGTTTGAGTACCG  310
             ||||||||||||||||||||
Sbjct  345   GCATGTCTGTTTGAGTACCG  364
```

(c) Alignment of GI: 194354257

(d) Alignment of GI: 532165358

Figures 15a, 15b, 15c, and 15d



Amoeboaphelidium sp. PML-2014 isolate FD01 18S ribosomal RNA gene, partial sequence;
Sequence ID: gb|JX967274.1| Length: 4667 Number of Matches: 3

Range 1: 3205 to 3622 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
431 bits(233)	2e-116	359/421(85%)	6/421(1%)	Plus/Plus

```
Query  639   CGATCTCAAATCAGACAAGACTACCCGCTGAACTTAAGCATATTAATAAGCGGAGGAAAA  698
             |||||||||||||||||||| |||||||||||||||||||||| ||||||||||||||||
Sbjct  3205  CGATCTCAAATCAGACAAGATTACCCGCTGAACTTAAGCATATYAATAAGCGGAGGAAAA  3264

Query  699   GAAACCAACAGGGATTCCCCCAGTAATGGCGAATGAAGCGGGAATAGCTCAAATTTTTAA  758
             ||||| |||| ||||||||  ||||| ||||| ||||| ||||| ||||||||||| |||
Sbjct  3265  GAAACTAACAAGGATTCCCATAGTAACGGCGAGTGAAGTGGGAACAGCTCAAATTTGTAA  3324

Query  759   TCTCTTCGGAGAGTTGTAATTTGAAGAGGTGACATCGTCGTCTTTGCCTGGTCAAAGTCT  818
             ||||||||||||||||||||||| ||||| |   ||| ||  ||  || |||  ||||||
Sbjct  3325  TCTCTTCGGAGAGTTGTAATTTGTAGAGGCGTTTTCGACG-GTTAACCGGGTAGAAGTCT  3383

Query  819   CCTGGAAAGGAGCAACATGGAGGGTGAAATTCCCGTATC-CGA-CCAGGTTGAAGGC-GC  875
             | ||| || ||||  ||  |||||||| | |||||| ||  || || |||  |  || | 
Sbjct  3384  CTTGGGAAAGAGCGTCACAGAGGGTGAGAATCCCGT-TCGTGATCCGGGTATACCGCAGA  3442

Query  876   TCTTGATTCATTCTCAAAGAGTCGGGTTGCTTGAGACTGCAGCCCAAAGTGGGTGGTATA  935
             |  |||| |  | |||||||||||||||| ||| ||||||||||| || | |||||||||
Sbjct  3443  T-ATGATACGCTTTCAAAGAGTCGGGTTGTTTGGGACTGCAGCCCTAAATTGGTGGTATA  3501

Query  936   TTCCATCTAAAGCTAAATATTGGCGAGAGACCGATAGCAAACAAGTACCGTGAGGGAAAG  995
             |||||||||||||||||||  ||||||||||||||||| ||||||||| |||| ||||||
Sbjct  3502  TTCCATCTAAAGCTAAATACAGGCGAGAGACCGATAGCGAACAAGTACTGTGAAGGAAAG  3561

Query  996   ATGAAAAGAACTTTGAAAAGAGAGTTAAAAGTACGTGAAATTGCTAAAAGGGAAACGTTT  1055
             |||||||||||| |||| ||||||||||||||||||||||||||||||||||||||||||
Sbjct  3562  ATGAAAAGAACTCTGAAGAGAGAGTTAAAAGTACGTGAAATTGCTAAAAGGGAAACGTTT  3621

Query  1056  G  1056
             |
Sbjct  3622  G  3622
```

Figure 15e

(e) Alignment of GI: 532166006

Figure 15: Alignment results of the five most abundant fungal sequences to their highest scoring BLAST hits of known phylum level taxonomy.

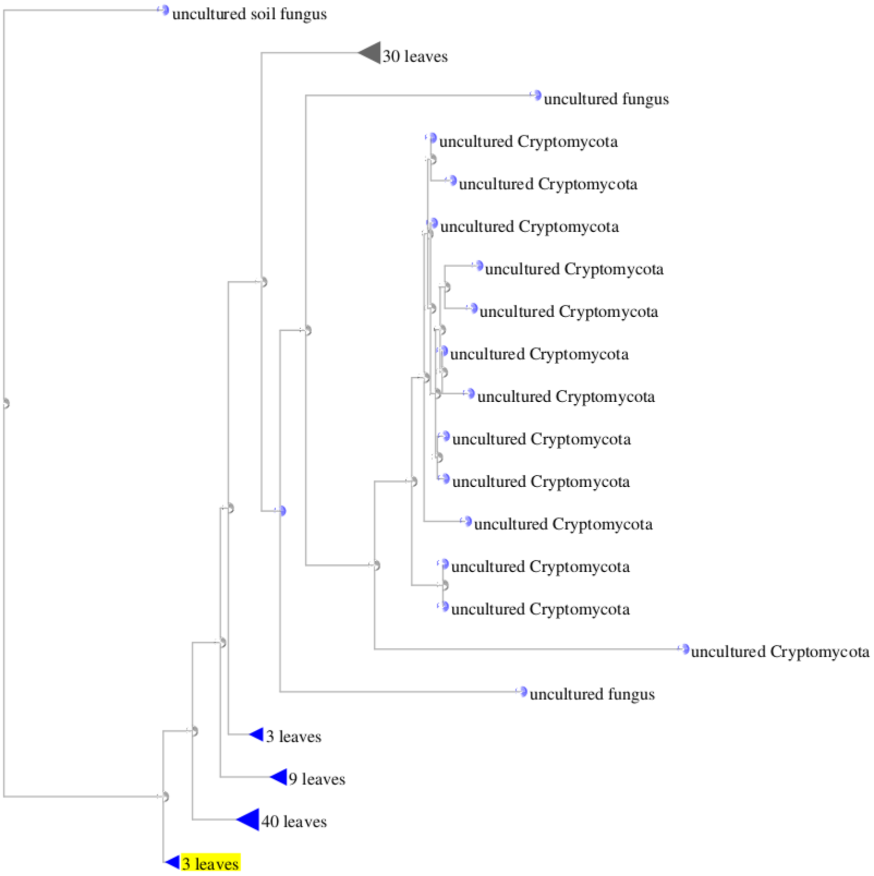


Figure 16: Distance tree for GI: 532165669, and GI: 532165968, collapsed on the branch highlighted in yellow.

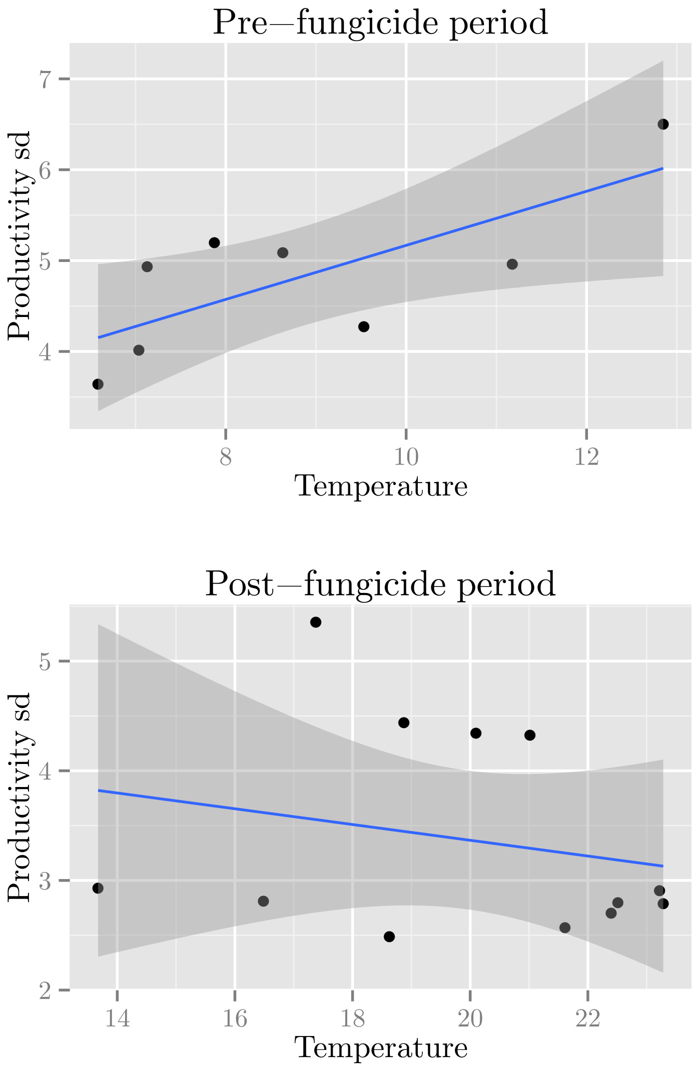


Figure 17: Pre- and post-fungicide temperature and productivity variability relationship.

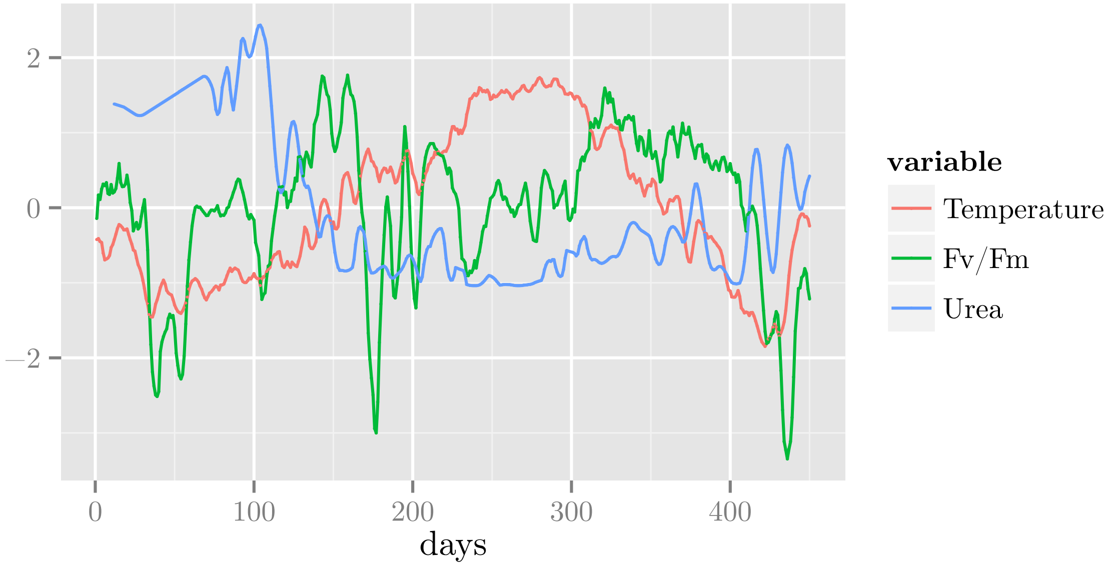


Figure 18: Select Phenotypes: Relationship of temperature, urea, and photosynthetic health (F_v/F_m) over time, standardised by centering around their mean and division by their standard deviation.

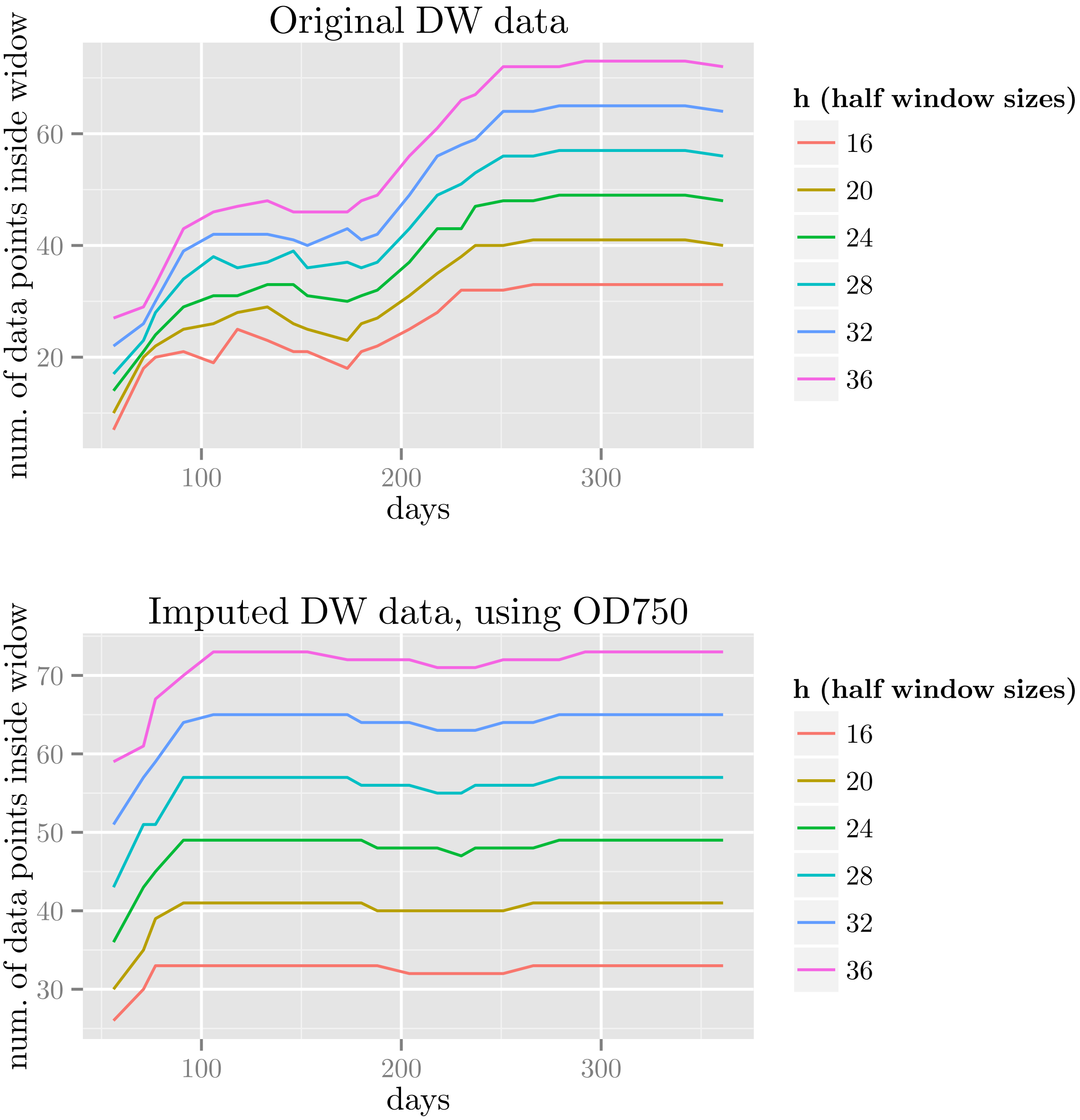


Figure 19: Number of available data points inside given half window (h) in original and imputed (using OD 750) DW (g/l) data.

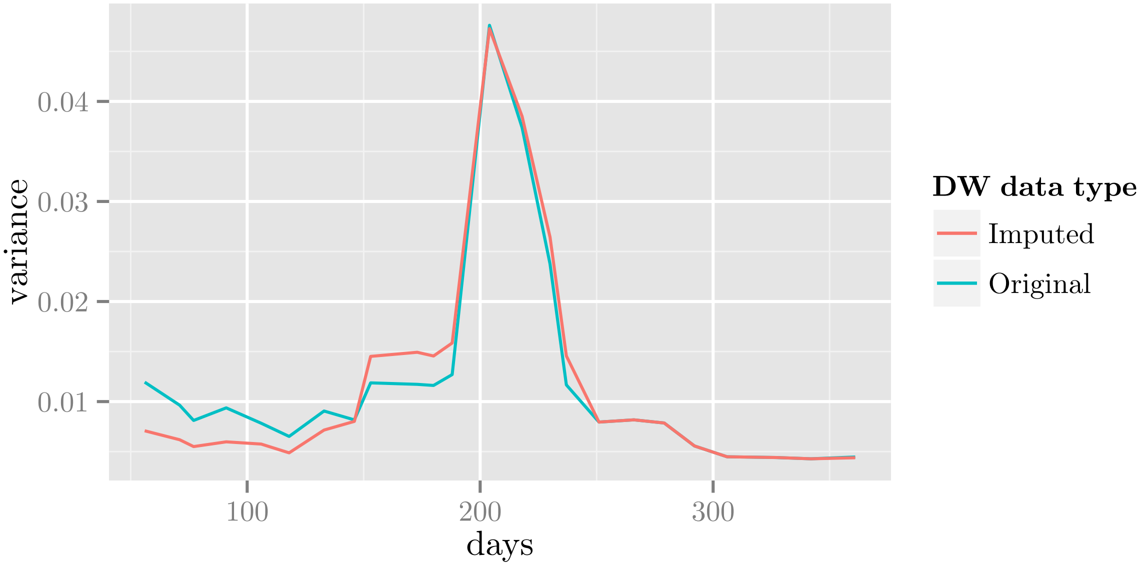


Figure 20: Variance patterns of original and imputed (using OD 750) DW (g/l) data using half window size of $h = 28$ days.

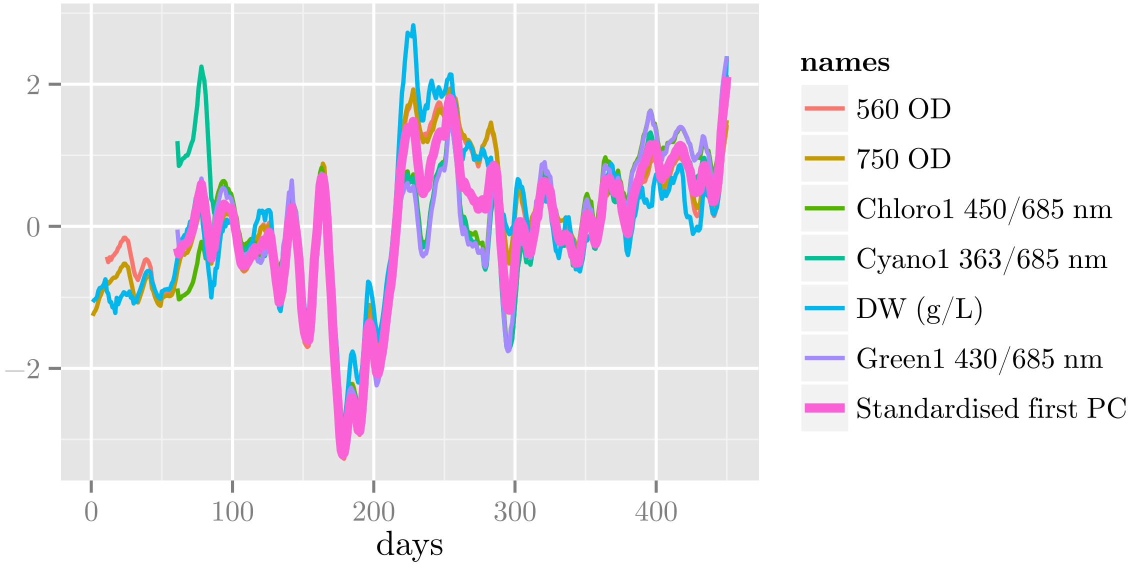
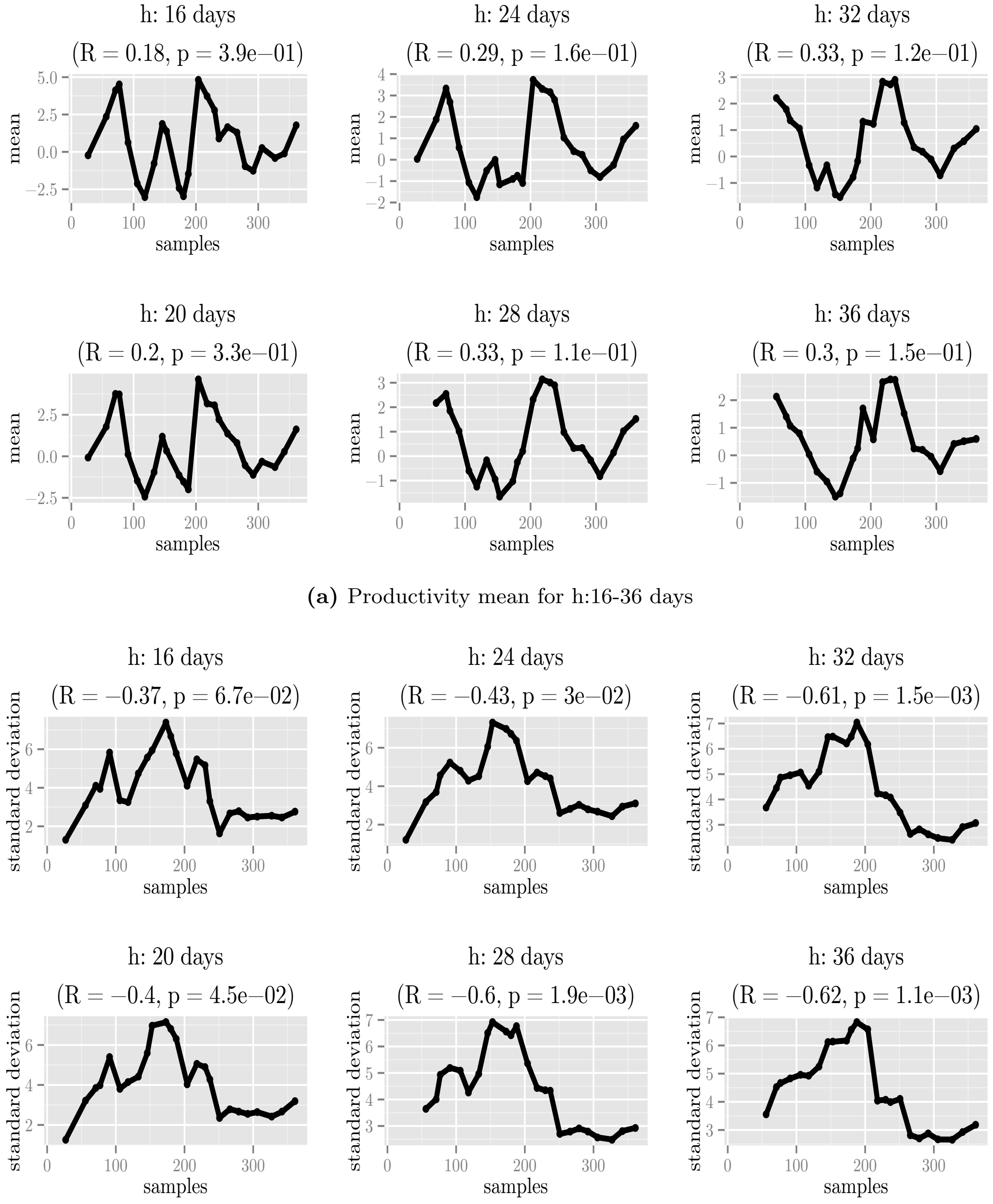


Figure 21: Example highly correlated phenotypic variable cluster: 7 phenotype variables (560 OD AVG, 750 OD AVG, DW g/L, Chloro1 450/685 nm AVG, Green1 430/685 nm AVG, KG, Cyano1 383/685 nm AVG) that mainly consist of various fluorescence levels and dry weight measures. Normalized variables, together with their first normalized principle component (dashed red), explaining 87.3% of the variance of the cluster.



(a) Productivity mean for h:16-36 days

(b) Productivity standard deviation for h:16-36 days

Figure 22: Productivity statistics trends for various h (half window) sizes changing from 16 to 36 days.