

Flow cytometry workflow for active, dormant, and dead microbial populations

M.L. Larsen

27 August, 2015

Overview

The microbes within communities have wide ranging metabolic activities that are influenced by resource availability. We can classify their metabolic activity with DNA and RNA content. Here, we seek to enumerate the fractions of the active (greater RNA than DNA), dormant (less RNA to DNA), and dead populations of microbes within freshwater ecosystems.

Document workflow

1. Setup work environment, load flow cytometry data for processing
2. Diagnostics with control data
3. Define static and dynamics gates
4. Process data files
5. University Lake Time Series Analysis
6. Indiana Ponds Analysis

Data Collection

Data for this project were collected as part of a long term sampling regime on University Lake in Bloomington, Indiana and wildlife refuge ponds scattered through south central Indiana. All data were collected in the Indiana University Flow Cytometry Core Facility operated by Christiane Hassel and processed by M.L. Larsen.

Data Processing

1. R Version and Package information

Source code used in this workflow were provided by R. Nikolic with addition source code written specifically for this project by M.L. Larsen.

All analyses were completed with R version 3.2.1 (2015-06-18) and the following packages:

<i>Package</i>	<i>Version</i>	<i>Module Description</i>
flowCore	1.34.9	data gating
flowQ	1.28.0	flow data quality assurance
flowViz	1.32.0	visualization of flow data

1. Load data

The data for this project were collected in the IU Flow Cytometry Core Facility (C. Hassel) using eFluor fixed viability dye, Hoescht 33342, and Pyronin Y.

```
## A flowSet with 7 experiments.  
##  
##   column names:  
##   FSC-A SSC-A FSC PMT-A Alexa Fluor 488-A PI (B)-A Pacific Blue-A APC-A
```

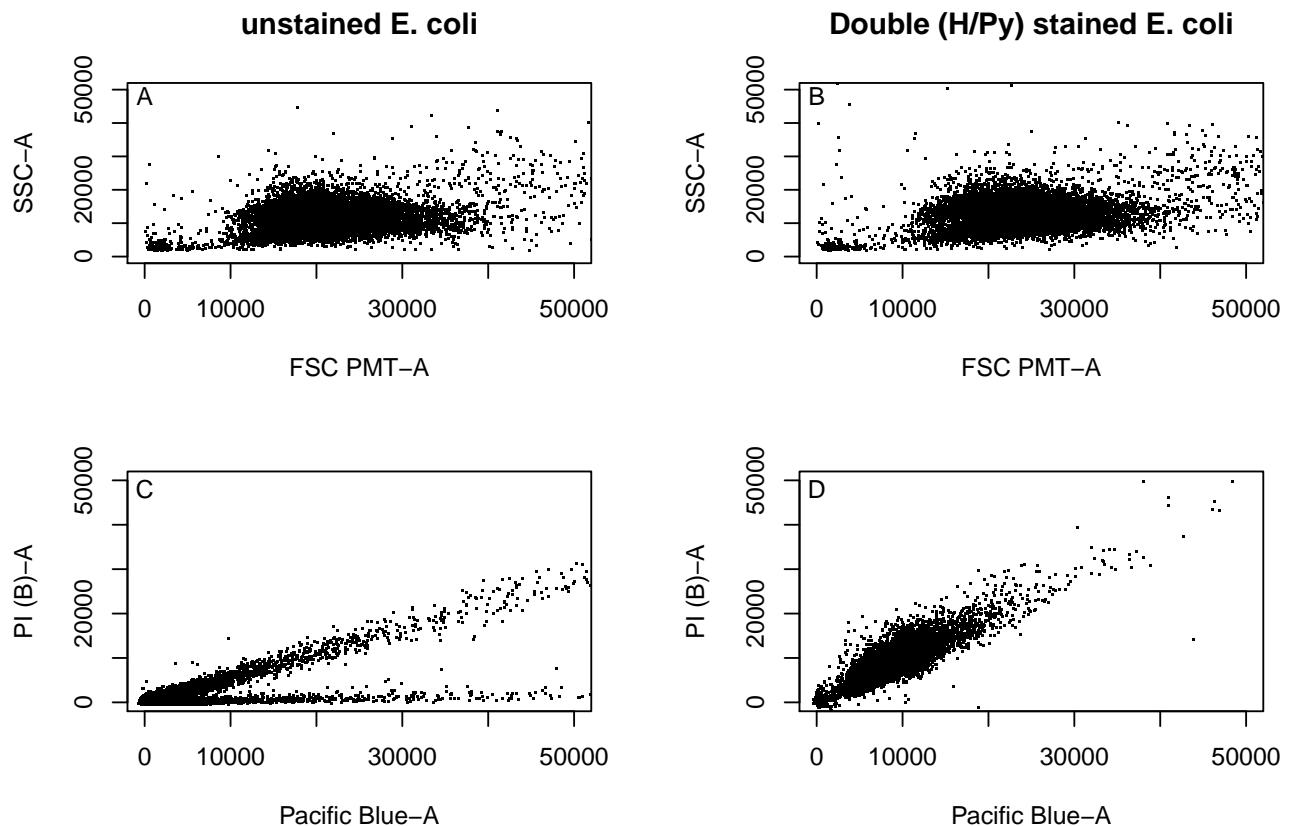
This experiment contains 7 samples.

2. Diagnostics with control data

2.1 Initial Visualization

Diagnostically evaluating the control data can help establish instrument malfunction, poor run quality, or the need for data preprocessing (e.g. compensations, transformation, etc.). `flowViz` provides the functionality to visually evaluate the data while `f1owQ` analyzes data quality assurance.

Bivariate density plots between stained and unstained samples show shifts in the data clustering (Figure 3 B,D). The *E. coli* standards in our control batch show very similar distributions in the FSC v SSC plots, however, the distribution changes substantially with the Pacific Blue v PI plot because of the presence of the stain in the cells. We should expect a diagonal shift up in the stained group.



2.2 Preprocessing

2.2.1 Compensation

Compensation is a normalization step by which excess fluorescence from the events read is removed.

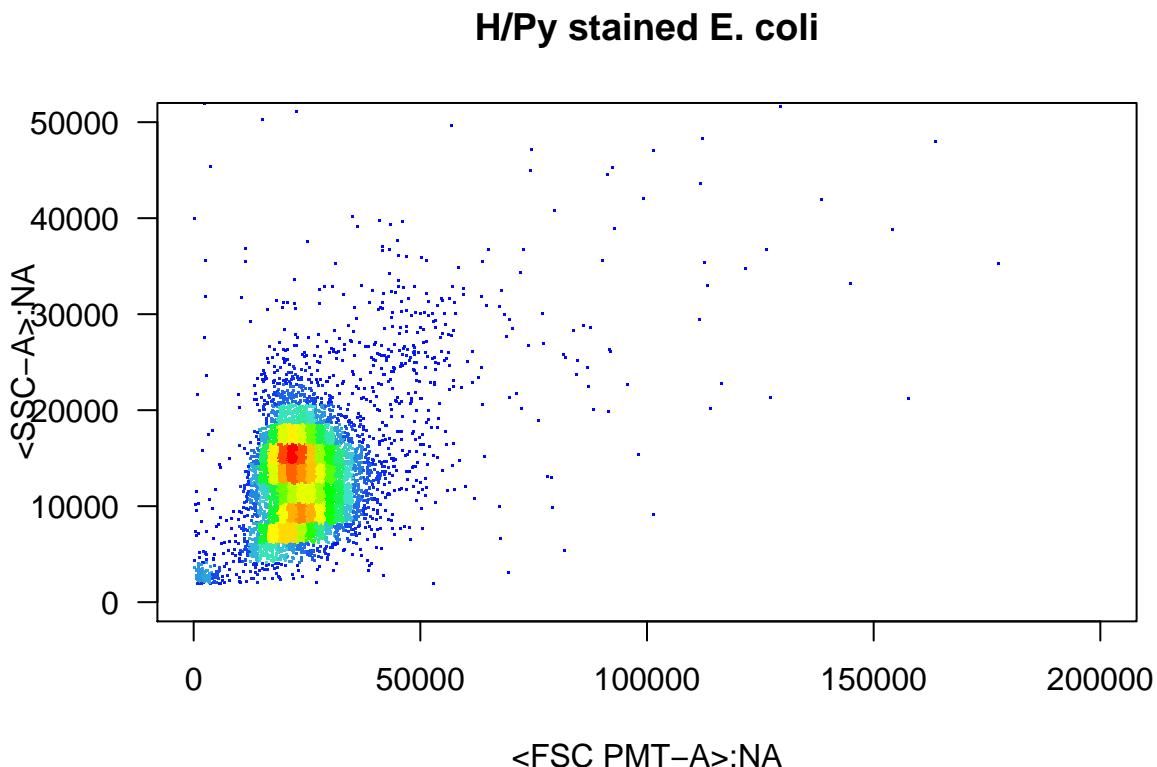
```
fs.controls[[1]]@description$'SPILL'

fs.controls.comp <- fsApply(fs.controls,function(frame){
  #extract compensation matrix from keywords
  comp <- keyword(frame)$`SPILL`
  new_frame <- compensate(frame,comp)
  new_frame
})

fs.controls.comp
summary(fs.controls.comp)
```

2.2.2 Visualize and Remove Margin events

Boundary, or marginal events, are common in the generation of flow cytometry data and should be considered as noise. These are events that fall out of the dynamic range that the flow cytometer can detect a signal. Before moving forward with the transformation, we must first classify and remove these boundary events with `flowQ` and `qaProcess.marginevents`, `qaProcess.cellnumber`, `qaProcess.timeline` and `qaProcess.timeflow`.

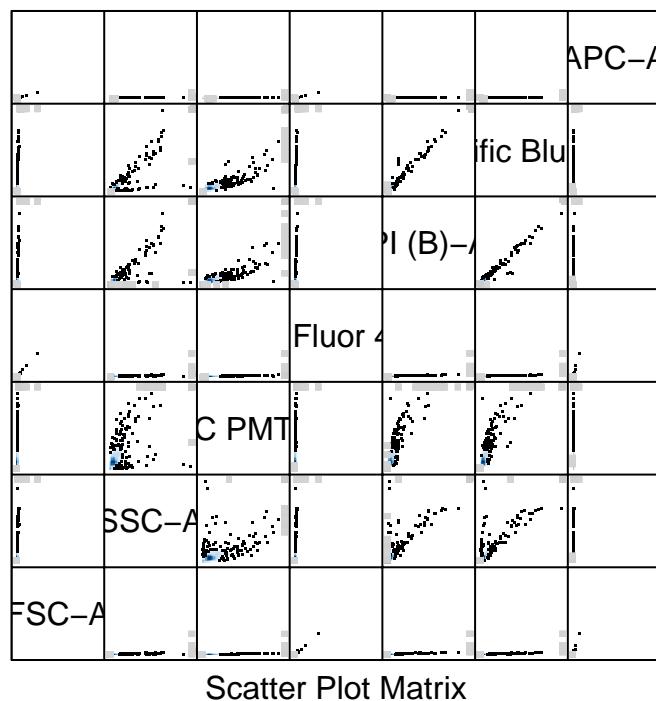


2.2.2.1 Time Anomalies `qaProcess.timeflow` & `qaProcess.timeline` can be used to detect disturbances in flow over time. Good even, flow through the machine shows that the instrument is running smoothly. These timeflow plots visualize acquisition rate over time, so the number of events that are recorded in a given time interval. The summary QA report for all the data in a `flowSet` can be found in the output.

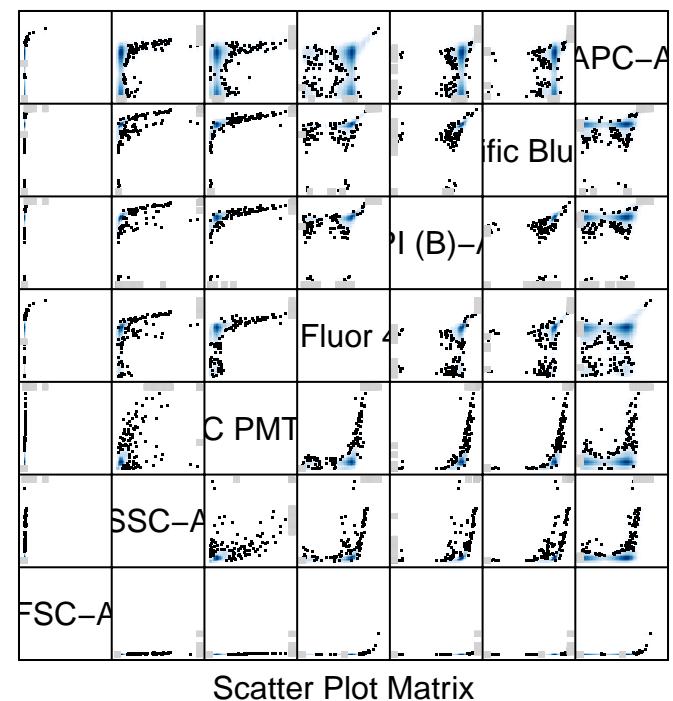
Based on the QA analysis, this flowset is ready for to be checked for transformation. However, there may be some cases where the marginal events need to be removed and/or the data normalized.

2.2.3 Transformation

Data transformations can be done in-line or out-of-line techniques. Using the compensated data, For more information about choosing the appropriate transformation, see Spidlen et al. 2006.



Scatter Plot Matrix



Scatter Plot Matrix

3. Define dynmanic gates using `flowDensity`

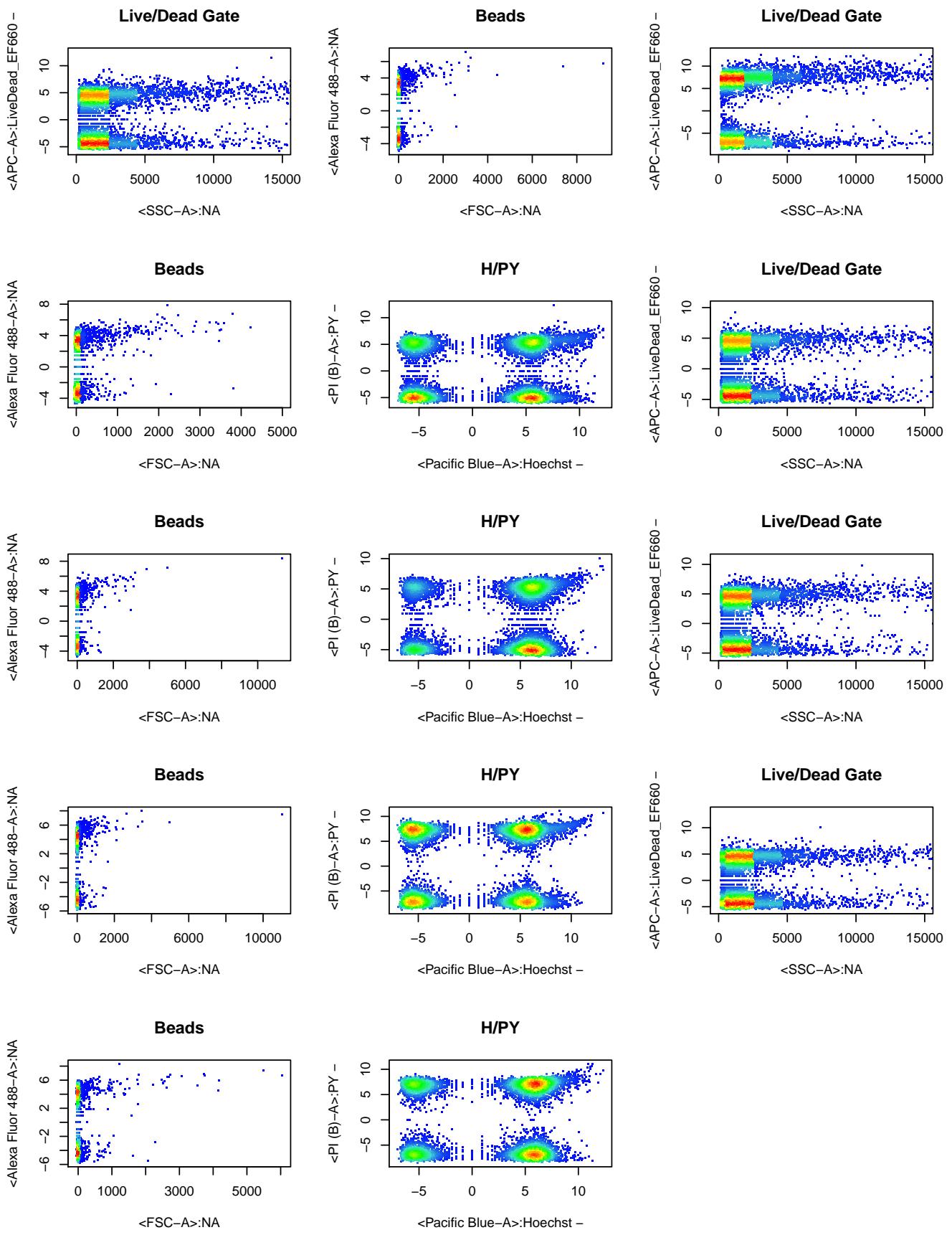
A flow cytometry `workFlow` object uses filters, or gates, that are related to the fluorescence of the eFluor 660 (a fixed viability dye), Hoechst 33342, or Pyronin-Y. This process uses all of the information to construct informational sorting for the data sets.

<i>Stain</i>	<i>Activity</i>	<i>EX/EM</i>
Molecular Probes eFluor 660	fixed viability dye	
Hoechst 33342	DNA specific binding	
Pyronin Y	RNA specific binding	

For each of the single color controls, we will use static gating to create the initial gate sets. From here, dynamic gates will using `kmeansFilter` will be applied to track data clusters. After constructing filters,

we can then count populations of events.

Let's look at the data in their respective channels. What's immediately obvious is that each of the controls look very similar! This will be corrected with the addition of the appropriate gates.



3.1 Create static gates

Creating a gate for the beads produces a standardized count for estimating population density (events/mL). Beads are most easily seen in the FSC-A v. Alexa Fluor 488.

```
# Create a gate for each single, double, or triple color control as well as beads
# background

bead.gate <- rectangleGate(filterId = "beads",
                            "FSC-A" = c(0,12), "Alexa-A" = c(0,12))
#live.gate <- kmeansFilter("APC-A" = c("Live", "Dead"), filterID = "live")
live.gate <- kmeansFilter("APC-A" = c(-10,0), "SSC-A" = c(0,12),
                           filterId = "live")

actdorm.gate <- kmeansFilter("Pacific Blue-A" = c("Active", "Dormant"), filterID="AD")
```

Use of the Live/Dead fixed viability dye requires that we use a negative gating process. Cells with compromised membranes will take up the stain while those with intact membranes will not.

4. Bring it all together now!

Active populations are cells that have RNA > DNA **Dormant** population have RNA = DNA **Dead** population is filtered out with the fixed viability dye

Create data file that contains information about the samples as well as calculations for each population following gates

4.1 Read in information for each sample

5. University Lake Time Series analysis

5.1 Temporal trends

With the data file created previously, we now have a way to look at the temporal trends associated with the microbial community in University Lake.

6. Indiana Ponds Geographic Abundances

6.1 Population relative abundances