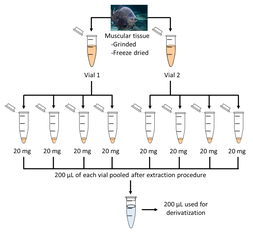
Implementation of Fisher ratio analysis for metabolite discovery in pacu fish using comprehensive two-dimensional gas chromatography with mass spectrometry

Sonia Schöneich[1], Sarah E. Prebihalo[1], Celina Monzòn[2], Robert E. Synovec[1]
[1]Department of Chemistry, University of Washington, Seattle WA
[2]Universidad Nacional del Nordeste, Corrientes, Argentina

Introduction

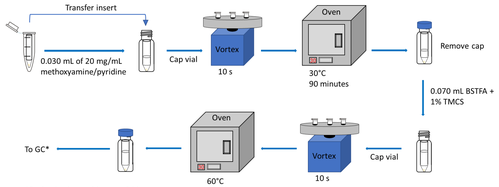
Tile-based Fisher ratio (F-ratio) analysis is used to discover and quantify 32 metabolites in Argentinian pacu fish, using data collected by comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry (GC×GC-TOFMS) via a LECO Pegasus BT 4D instrument. The experimental design for Fisher ratio (F-ratio) analysis has two sample classes: pacu fish (class one), and pacu fish spiked with a 32 metabolite standard (class two). Pacu fish are often farmed along with rice crops, which introduces sources of contamination that may alter the fish metabolome, making such a study challenging. To address this challenge, this GC×GC-TOFMS instrument provides outstanding separation peak capacity and detection sensitivity. Workflow principles to optimally apply tile-based F-ratio analysis are also presented.

Sample preparation

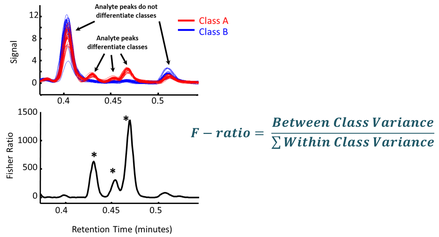


Above: Muscle tissue from the Argentinian pacu fish was grinded, freeze dried, and split between two vials for storage. In order to account for fish & vial heterogeneity, four extracts were prepared from vial 1, and four from vial 2, which were pooled prior to derivatization.

Below: Due to the nature of metabolites, derivatization after extraction is necessary prior to GC analysis. For each sample, derivatization was performed via a two-step process, as illustrated below: First, oximation was performed with the addition of 300 µL of a methoxyamine/pyridine solution followed by heat for 90 minutes. Second, trimethylsilylation was employed using 700 µL of N,O,-Bis(trimethylsilyl)trifluoroacetamide (BSTFA) + trimethylsilyl chloride (TMCS).



F-ratio

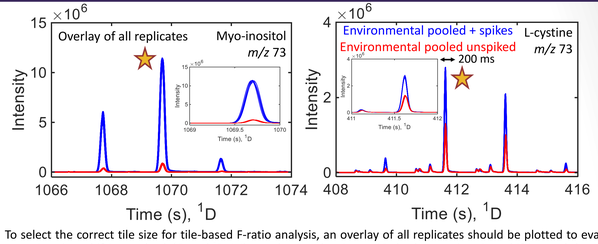


$$F-ratio = \frac{Between\ Class\ Variance}{\sum Within\ Class\ Variance}$$

F-ratio analysis is characterized as a supervised, non-targeted method, where the classes are known but the components that distinguish between classes are unknown. Notably, F-ratio emphasizes true class-distinguishing ability over absolute signal. F-ratio values can range from zero (little variance between classes), and infinity, where the magnitude scales with the extent of between class variance relative to within-class variance.

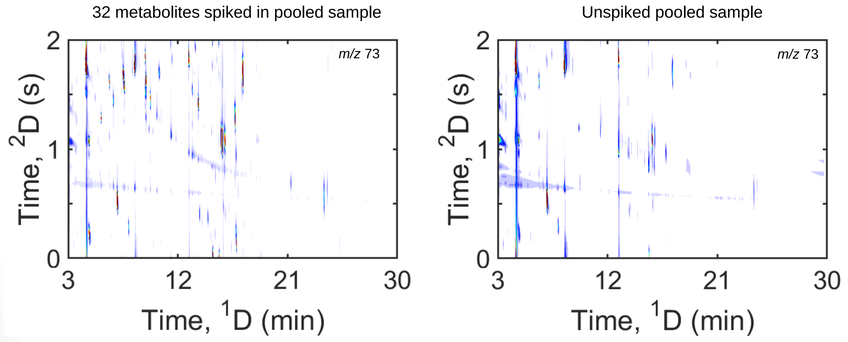
Determining tile size

Examples of analytes with native traces of the spiked metabolites. Left: The myo-inositol peak for each spiked (blue) and unspiked (red) replicate showing there is essentially no t_R shifting on either dimension. The insert is provided for better visualization of the lack of retention time shifting. Right: L-cystine peak for the spiked (blue) and unspiked (red) representative for the narrower peaks with ^2W_b=200 ms. The inset is a zoom-in of the tallest modulated peak provided for visualization purposes.



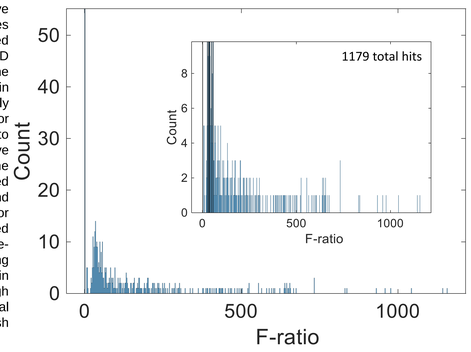
To select the correct tile size for tile-based F-ratio analysis, an overlay of all replicates should be plotted to evaluate the retention time shifting on both ^{1}D and ^{2}D. When selecting the size on ^{1}D, both retention time shifting and the number of modulations should be taken into account. The size of the ^{2}D tile size should be selected as to account for the width of the modulated ^{2}D peaklets as well as any retention time shift. A tile that is too small may result in many redundant hits or false positives in the hit list, while a tile size too large may not discover small analytes that may be in the same tile as another analyte with greater signal. In this study, each analyte is modulated 2-4 times, so a ^{1}D tile size of 6 modulations, or 12 s, will account for any shifts on ^{1}D. The ^2W_b ranged from 200 ms to 400 ms, which supports setting the ^{2}D tile size to 400 ms. For this data set, the selected tile size on ^{1}D was 12 seconds (6 modulations) and 400 ms on ^{2}D.

Spiked and unspiked fish samples



GC×GC-TOFMS chromatograms of the selective *m/z* 73 for the derivatized pacu fish samples spiked with 32 metabolites (left) and unspiked (right) collected with the LECO Pegasus BT4D (LECO Corportation, St. Joseph, MI). The metabolites spiked undergo trimethylsililation in the derivatization process and are therefore easily discovered using the selective *m/z* 73 for trimethylsilyl groups. In order for F-ratio to discover the spiked metabolites, the native analytes should have the same signal in both the unspiked and spiked samples. This was ensured by diluting the native analytes in both spiked and unspiked samples by a factor of 3 (50 µL spike or pyridine:100 µL sample). These spiked metabolites should be easily discovered using tile-based F-ratio that finds class distinguishing features between sample classes. Difficulties in discovering the analytes may be due to high within class variance, small differences in signal between classes, innate differences between fish samples, and choosing the incorrect tile size.

F-ratio hit list for metabolite discovery



The F-ratio hit list was generated using the tile size of 12 s (6 modulations) on ^{1}D and 400 ms on ^{2}D. A total of 22 metabolite spikes were discovered in the top 32 hits with the hit list generating a total of 1179 hits (distribution to the right). The hit list was prepared after mining and removing spurious hits including trimethyl silanol and various trimethylsilylated phosphoric acid fragments. To reduce the number of redundant hits, the cluster window was 8 s on ^{1}D and 200 ms on ^{2}D and the number of *m/z* required to pass the signal-to-noise (*S/N*) threshold of 50 was set to 10 *m/z*. The F-ratio is the average F-ratio of the top 3 *m/z*. The metabolites that appear higher on the hit list are most likely those with higher signal in the standard mixture and have significantly different signals relative to the unspiked fish samples. Some analytes may fall further down the list due to greater within class variance, while others may be native to the unspiked samples. Retention times are reported and retention time shifts relative to the standard mixture were calculated as $\Delta^n t_R = t_{R,spiked} - t_{R,std}$.

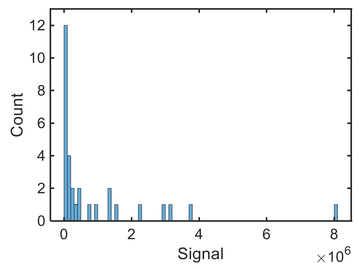
Hit #	ID	F-ratio	^1t_R (s)	^2t_R (s)	Δ^1t_R (s)	Δ^2t_R (ms)
1	Fumaric acid	1047	590	1.84	2	-70
4	Malonic acid	931	488	1.57	2	-30
5	Pyroglutamic acid	837	850	1.42	2	-10
6	L-glutamic acid	831	822	1.54	2	-40
7	Uracil	792	646	0.68	2	-850
8	Maleic acid	703	614	1.47	0	-40
10	4-methylvaleric acid	676	588	1.29	0	-140
11	Succinic acid	667	594	1.62	2	-40
12	Myo-inositol	649	1068	1.71	0	-70
13	Glycerol-3-phosphate	646	920	0.03	0	-40
14	DL-lactic acid	645	660	1.68	4	-70
15	Gluconic acid	642	1036	0.14	0	-60
16	D-glucose-6-phosphate	640	1032	1.3	0	-60
17	D-galactose	637	964	1.21	2	50
18	Oxalic acid	596	434	1.43	4	0
19	D-glucose	581	970	1.07	2	-70
20	L-cystine	503	412	1.6	2	-50
21	Maltose	492	1470	0.53	-2	-70
22	Methylmalonic acid	464	480	1.81	2	110
24	Pyruvate	321	372	1.28	2	10
26	L-alanine	286	326	1.14	4	20
30	Glycine	229	358	1.05	0	-850

50 ppm metabolite standard mix

ID	^1t_R (s)	^2t_R (s)	Signal
D-glucose-6-phosphate	1032	1.4	163456
Methylmalonic acid	478	1.72	2256384
Oxalic acid	430	1.43	248832
Myo-inositol	1068	1.78	905600
Malonic acid	486	1.6	1581056
Pyruvate	370	1.27	916480
Maltose	1472	0.6	18352
Aspartic acid	820	1.11	13240
Succinic acid	592	1.66	3726848
Glycerol-3-phosphate	920	0.07	376960
L-glutamic acid	820	1.58	58936
L-lysine	828	1.51	10816
L-alanine	322	1.12	809696
L-methionine	720	1.83	22432
DL-lactic acid	656	1.75	424448
Isovaleric acid	618	1.67	59288

ID	^1t_R (s)	^2t_R (s)	Signal
Uracil	648	1.53	179440
Creatinine	802	1.88	590336
Pyroglutamic acid	848	1.43	681968
Myristic acid	1290	0.15	8688
Maleic acid	614	1.51	1334784
Glutamic acid	914	1.17	174400
Fumaric acid	588	1.91	8325568
Gluconic acid	1036	0.2	163456
D-glucose	968	1.14	273536
D-galactose	962	1.16	417728
L-serine	526	1.73	13944
L-threonine	370	1.27	11704
L-cystine	410	1.65	3108864
L-proline	938	1.09	9776
4-methylvaleric acid	626	1.43	73216
Glycine	358	1.9	46656

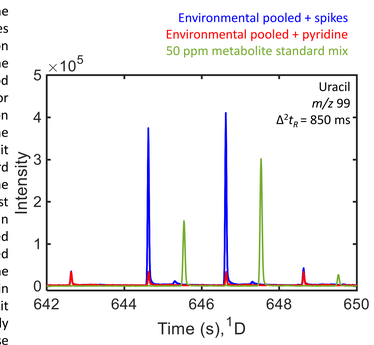
A standard mixture of 32 metabolites was used to generate an in-house library with retention times and signal of the tallest modulation for each analyte. These 32 metabolites were later spiked into pacu fish samples for analyte discovery by tile-based F-ratio analysis and quantification. Retention times will be used to mine the hit list for the spikes and signals will be used for quantification.



The signals in the last column of the tables to the left were used to generate the histogram above. The 12 analytes with the lowest signal were the most difficult to identify reliably and are expected to be the most difficult analytes to discover by tile-based F-ratio analysis by the spikes alone. It is expected that the other 20 analytes with higher signal will be higher on the F-ratio hit list.

Metabolite quantification

F-ratio analysis provides the benefit of finding analytes with retention time shift on either ^{1}D or ^{2}D. While the standard addition method is a popular method for quantification, it relies on analyte spikes to retain the same retention time as it would in the neat standard mix. The example to the right displays the highest signal *m/z* 99 of uracil in the spiked and unspiked fish samples that is shifted by 850 ms relative to the modulated ^{2}D uracil peak in the standard mix. This hit would not have been easily discovered without the use of F-ratio.



After analyte discovery by tile-based F-ratio analysis, the spiked metabolites in the top 50 hits were quantified using the signal for the tallest modulated peak. This was systematically done using the pin locations from the F-ratio hit list for all replicates. The standard addition method was used to determine the concentration of the metabolites native to the fish was in parts per million (ppm) and then multiplied by the dilution factor. Hits higher on the list were in general lower in concentration in the unspiked samples.

Hit #	ID	Concentration (ppm)
1	Fumaric acid	0.90
4	Malonic acid	0.09
5	Pyroglutamic acid	1.35
6	L-glutamic acid	1.55
7	Uracil	37.10
8	Maleic acid	1.04
10	4-methylvaleric acid	5.02
11	Succinic acid	0.42
12	Myo-inositol	13.30
13	Glycerol-3-phosphate	0.33
14	DL-lactic acid	1.56
15	Gluconic acid	2.36
16	D-glucose-6-phosphate	1.22
17	D-galactose	17.43
18	Oxalic acid	1.05
19	D-glucose	39.99
20	L-cystine	79.72
21	Maltose	20.19
22	Methylmalonic acid	0.08
24	Pyruvate	0.64
26	L-alanine	132.58
30	Glycine	19.52

References

- L.C. Marney, W.C. Siegler, B.A. Parsons, J.C. Hoggard, B.W. Wright, R.E. Synovec, *Talanta*. 2013, 115, 887–895.
- B.A. Parsons, L.C. Marney, W.C. Siegler, J.C. Hoggard, B.W. Wright, R.E. Synovec. *Anal. Chem*. 2015, 87 (7), 3812–3819.
- B.A. Parsons, D.K. Pinkerton, B.W. Wright, R.E. Synovec, *J. Chromatogr. A*. 2016, 1440, 179-190.
- N.E. Watson, B.A. Parsons, R.E. Synovec, *J. Chromatrogr. A*. 2016, 1459, 101-111.
- R.C. Reaser, B.W. Wright, R.E. Synovec, *Anal. Chem*. 2017, 89, 3606-3612.