

Supplementary materials

A Coding Basis and Three-in-One Integrated Data Visualization

Method ‘Ana’ for the Rapid Analysis of Multidimensional Omics Dataset

Hefei Zhao and Selina C. Wang *

Department of Food Science and Technology, University of California, Davis, One Shields Ave,
Davis, CA 95616, USA

*Corresponding: scwang@ucdavis.edu

S. Codes

S. Code 1 Save the following codes between the two lines in the 'Ana.m' file in MATLAB as shown in S. Fig 2

%-----

% 09/02/2022 the .m file 'Ana' version 1.0 was created by Hefei Zhao, PhD. This code is designed for analyzing olive pomace polyphenols concentrations from HPLC-DAD as well as many other omics data.

% Supplementary materials. S. Code 1

% A coding basis and three-in-one integrated data visualization method 'Ana' for the rapid analysis of multidimensional omics dataset

% Hefei Zhao¹, Selina C. Wang^{1,*}

% ¹Department of Food Science and Technology, University of California, Davis, One Shields Ave, Davis, CA 95616, USA

% *Corresponding author: Selina C. Wang, email: scwang@ucdavis.edu

% First author: Hefei Zhao, email: hzhao@huskers.unl.edu; hefzhao@ucdavis.edu

%

% References

% Zhao, H., Avena-Bustillos, R. J., & Wang, S. C. (2022). Extraction, Purification and In Vitro Antioxidant Activity Evaluation of Phenolic Compounds in California Olive Pomace. In Foods (Vol. 11, Issue 2).

<https://doi.org/10.3390/foods11020174>

% matlab - How I obtain bars with function bar3 and different widths for each bar? - Stack Overflow.

(n.d.). Retrieved September 3, 2022, from <https://stackoverflow.com/questions/24269516/how-i-obtain-bars-with-function-bar3-and-different-widths-for-each-bar>

% %%%%%%%%%%

% The percentage icon '%' means the contents after the % are all text notes instead of executable code.

clear, clc% clean RAM and command window

close all% close all figures, but Clustergram 1 must be closed manually.

tic % start timing

% %%%%%%%%%%

% Initial input area for normal users or beginners

fn= 'olivephenolics';% input the excel file name in the " area, olivephenolics can be replaced by user's excel file name, the excel file must be in the same folder as the this .m file

unt= 'mg/g';% input unit of data in the " area, mg/g can be replaced by the user's unit, such as %, g/mL, mg/mL etc.

fs= 20;% input font size, 18-22 are recommended, must be >=7

cl= jet(256);% color style: jet(256) is rainbow; cool is blue to pink; parula is blue to yellow; redbluecmap is blue to red; [] is transparent; user can replaced jet(256) by cool, [] or parula or bredbluecmap, etc.

pcamz= 10;% PCA marker size

pcalable= 0;% 0 will label PCA vectors by variable/ compound numbers; 1 will label PCA vectors by variable/ compound full names.

mk= '.'; % set PCA data marker in " area; . is dot; p is star; s is square; * is snow flasker; o is o; 'd' is rhombus.

% Note: Method for print or output figures:

```
% Figure 1, 3D heatmap
% print(fig3Dbar,'olive-adjustsize.png','-dpng','-r150');% -r150 defines 150dpi,-r300 will provide 300 dpi, -
r100 will provide 100 dpi, etc. See S. Fig. 3
```

% Clustergram 1, heatmap cluster, can only be output to a pdf for high-resolution as described in S. Fig. 4.

```
% Figure 2-7, print PCA charts automatically at 150 dpi
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
% It may result in errors if modify code below this line if not skillful on MATLAB.
```

```
% But it would be greatly helpful for active learners to modify the code to improve coding skills.
```

```
fname= strcat(fn);% excel file name
info= readtable(fname,'PreserveVariableNames',true, 'ReadRowNames',true);% read data again for
column names
rm= string(info.Properties.RowNames);% read row names of samples
cm= string(info.Properties.VariableNames);% read column/variable names of phenolic compounds
sz= size(info);% read data area size
ns= sz(1);% read sample number
nb= sz(2);% read compound number
num= readmatrix(fname); % read data only
num(:,1) = []; % delete first column NAN
ave= num; % define ave values
%

% plot 3D bar chart with size adjustments
%
fig3Dbar=figure; % name figure as fig3Dbar
h=bar3(ave,'detached');% 'detached', 'grouped', or 'stacked'
%
values=ave; % values equal to ave.
m= 1.1*max(values(:))*2; % normalize constant for bar width, 1.1 ensure a small distance between bars
shading interp % set color shading properties
for i = 1:length(h)% the following code set bottom size in accordance with the z-axis values
    % Get the ZData matrix of the current group
    xdata= get(h(i),'Xdata');
    ydata= get(h(i),'Ydata');
    zdata= get(h(i),'Zdata');
    set(h(i),'Cdata',zdata)
    for k = 1:size(xdata,1)/6
        datax = xdata((k-1)*6+1+(0:5),:);
        datay = ydata((k-1)*6+1+(0:5),:);
        dirx=((datax-round(datax))>0)-((datax-round(datax))<0);
        diry=((datay-round(datay))>0)-((datay-round(datay))<0);
        xdata((k-1)*6+1+(0:5),:) = round(xdata((k-1)*6+1+(0:5),:),0)+dirx*values(ceil(((k-1)*6+1)/6),i)/m;
        ydata((k-1)*6+1+(0:5),:) = round(ydata((k-1)*6+1+(0:5),:),0)+diry*values(ceil(((k-1)*6+1)/6),i)/m;
    end
end
```

```

    set(h(i),'XData',xdata);
    set(h(i),'YData',ydata);
end
set(h,'EdgeColor','k') % set edge color as k black
view(-25, 83); % default view angle
colormap (cl) % set colormap as cl
colorbar % show color bar
%
xticks(1:nb)% add x ticks
yticks(1:ns)% add y ticks
set(gca,'XTickLabel',cm)% label x axis by column/variable names of phenolic compound
set(gca,'YTickLabel',rm)% label y axis by row names of samples
set(h,'FaceAlpha',.5) % set transparency of bars to 0.5
xlabel(unt,'FontSize',fs) % set font size of z-axis
ax = gca;
ax.FontSize = fs; % set font size of color code bar
cb=colorbar;
colormap(cl); % define color as described by cl
cb.Label.String = unt; % set unit of color code bar as described by unt
%print(fig3Dbar,'olive-adjustsize.png','-dpng','-r150');% -r150 defines 150dpi,-r300 will provide 300 dpi, -
r100 will provide 100 dpi, etc.

% Cluster analysis
cfac=clustergram(ave, 'RowLabels',
rm,'ColumnLabels',cm,'Colormap',colormap(cl),'Standardize','Row'); % the code standardize data on each
row
% For more information see: https://www.mathworks.com/help/bioinfo/ref/clustergram.html
set(cfac,'Linkage','complete','Dendrogram',10)
set(cfac,'Annotate','off') % turn of annotate
set(cfac,'Linkage','Average') % set linkage method by Average
set(cfac,'RowPDist','Euclidean') % row distance method Euclidean
set(cfac,'ColumnPDist','Euclidean') % column distance method Euclidean
%
% PCA
sdz = zscore(ave,[],2); % standardized data along data rows
[coefs,score,latent,tsquared,explainedvariance] = pca(sdz); % run PCA
%
if pcalable== 0 % 0 will lable PCA vectors by variable/ compound numbers; 1 will lable PCA vectors by
variable/ compound full names.
    lbls= string(1:length(cm));
else
    lbls= cm;
end
%
figPCA12biplot= figure; % open a new figure
pa12= biplot(coefs(:,1:2),'Scores',score(:,1:2),'VarLabels',lbls,'Marker',mk,'MarkerSize',pcamz); % plot
PCA biplot of PC1 vs. PC2
grid off % turn off grid

```

```

box off % ture off box
ax = gca;
ax.FontSize = fs-6; % set font size of tick, 6 less than label font size
xlabel(['PC1', ' ', num2str(explainedvariance(1)), '%'], 'FontSize', fs) % label x-axis
ylabel(['PC2', ' ', num2str(explainedvariance(2)), '%'], 'FontSize', fs) % label y-axis
print(figPCA12biplot, 'PCA12-biplot.png', '-dpng', '-r150'); % print figure at 150 dpi
%
clr = hsv(ns);
figPCA12score= figure;
gscatter(score(:,1),score(:,2),rm,clr,mk) % plot PCA scoreplot of PC1 vs. PC2
legend('Location','northeastoutside')
grid off % turn off grid
box off % ture off box
ax = gca;
ax.FontSize = fs-6; % set font size of tick, 6 less than label font size
xlabel(['PC1', ' ', num2str(explainedvariance(1)), '%'], 'FontSize', fs) % label x-axis
ylabel(['PC2', ' ', num2str(explainedvariance(2)), '%'], 'FontSize', fs) % label y-axis
print(figPCA12score, 'PCA12-score.png', '-dpng', '-r150'); % print figure at 150 dpi
%
figPCA23biplot= figure; % open a new figure
pa23= biplot(coefs(:,2:3), 'Scores', score(:,2:3), 'VarLabels', lbls, 'Marker', mk, 'MarkerSize', pcamz); % plot
PCA biplot of PC2 vs. PC3
grid off % turn off grid
box off % ture off box
ax = gca;
ax.FontSize = fs-6; % set font size of tick, 6 less than label font size
xlabel(['PC2', ' ', num2str(explainedvariance(2)), '%'], 'FontSize', fs)
ylabel(['PC3', ' ', num2str(explainedvariance(3)), '%'], 'FontSize', fs)
print(figPCA23biplot, 'PCA23-biplot.png', '-dpng', '-r150');
%
figPCA12score= figure;
gscatter(score(:,2),score(:,3),rm,clr,mk) % plot PCA scoreplot of PC2 vs. PC3
legend('Location','northeastoutside')
grid off % turn off grid
box off % ture off box
ax = gca;
ax.FontSize = fs-6; % set font size of tick, 6 less than label font size
xlabel(['PC2', ' ', num2str(explainedvariance(2)), '%'], 'FontSize', fs) % label x-axis
ylabel(['PC3', ' ', num2str(explainedvariance(3)), '%'], 'FontSize', fs) % label y-axis
print(figPCA12score, 'PCA23-score.png', '-dpng', '-r150'); % print figure at 150 dpi
%
figPCA123biplot= figure; % open a new figure
pa123= biplot(coefs(:,1:3), 'Scores', score(:,1:3), 'VarLabels', lbls, 'Marker', mk, 'MarkerSize', pcamz); % plot
PCA biplot of PC1 vs. PC2 vs. PC3
grid off % turn off the grid
ax = gca;
ax.FontSize = fs-6; % set font size of tick, 6 less than label font size
xlabel(['PC1', ' ', num2str(explainedvariance(1)), '%'], 'FontSize', fs)

```

```

ylabel(['PC2',' ', num2str(explainedvariance(2)),'%'],'FontSize',fs)
xlabel(['PC3',' ', num2str(explainedvariance(3)),'%'],'FontSize',fs)% label z-axis
box on % turn on box
print(figPCA123biplot,'PCA123-biplot.png','-dpng','-r150');
%
ev= figure;
pareto(explainedvariance,0.99) % draw Explained Variance
ax = gca;
ax.FontSize = fs-6; % set font size of tick, 6 less than label font size
xlabel('Principal Component','FontSize',fs)
ylabel('Explained Variance(%)','FontSize',fs)
box off % turn off box
print(ev,'PCA-Explained-Variance.png','-dpng','-r150');
toc% stop timing

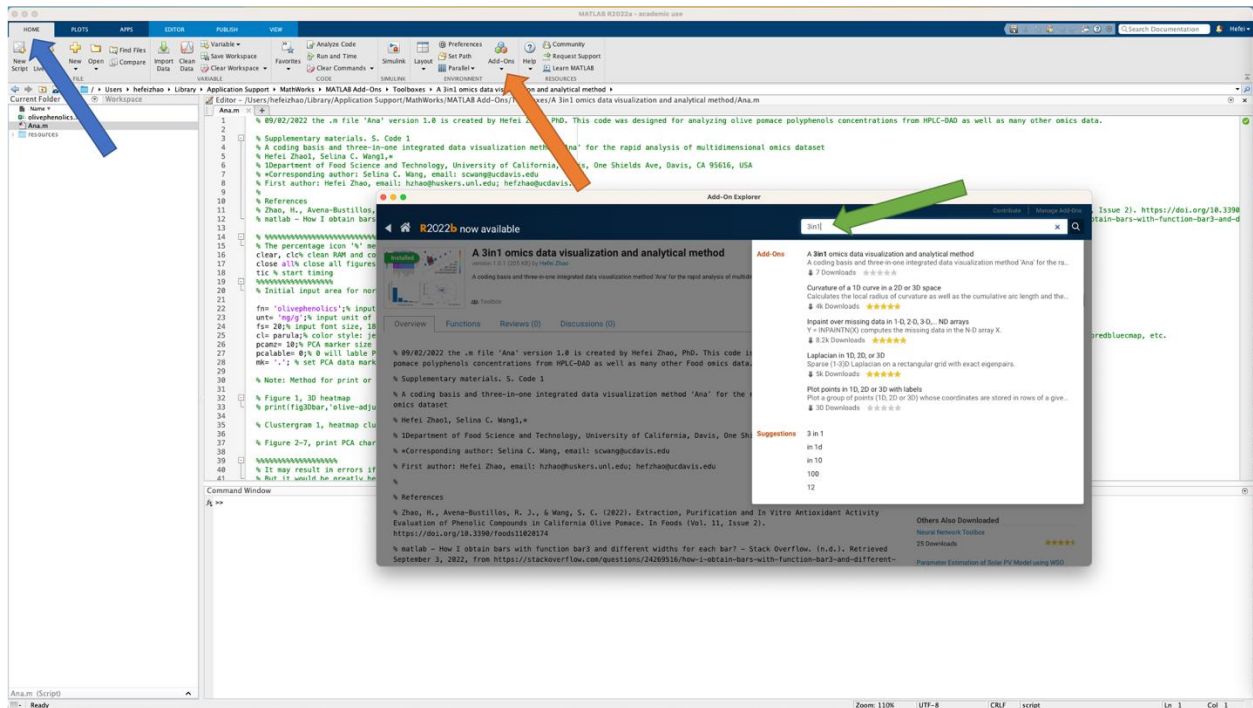
%-----

```

S. Figures

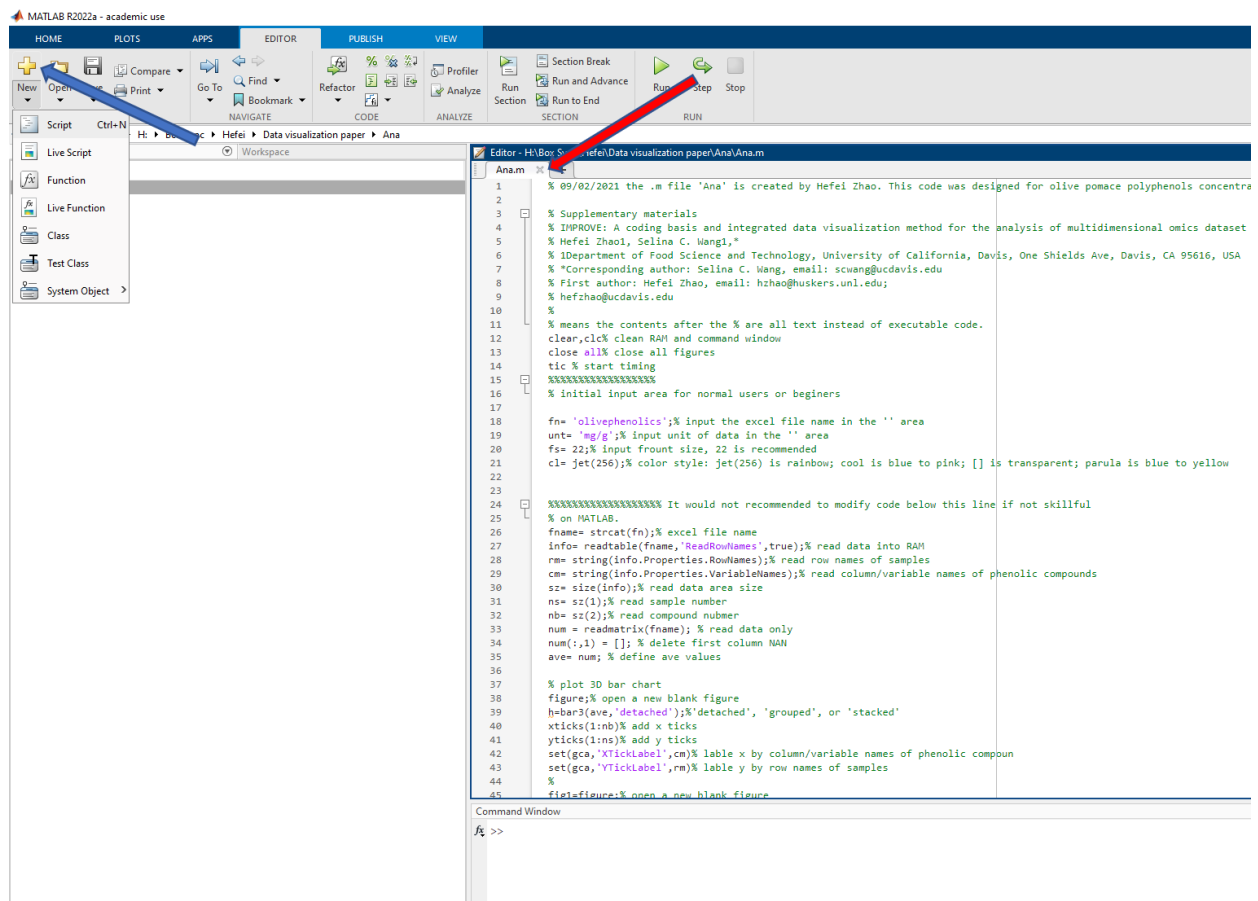
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	NAME	01.Vanillin	02.Gallic	03.Hydrox	04.Hydrox	05.Tyroso	06.Tyroso	07.4-HPA	08.Vanillin	09.Caffeic	10.Vanillin	11.p-coum	12.Ferulic	13.Verbas	14.Rutin	15.Luteoli	16.o-coum	17.Apigen	18.Oleuro	19.Pinorei	20.Cinnan	21.Luteoli	22.Apigen	
2	WE in DOP	0.165	0.223	1.407	1.978	1.096	0.460	1.700	0.203	0.050	0.371	0.084	0.047	0.833	0.770	0.042	0.101	0.055	0.811	0.084	0.027	0.010	0.007	
3	70M in DOP	0.054	0.045	0.657	2.017	0.679	0.365	0.800	0.208	0.044	0.329	0.097	0.023	1.074	1.360	0.042	0.070	0.121	1.270	0.257	0.019	0.487	0.062	
4	70E in DOP	0.225	0.000	0.250	1.356	0.384	0.162	0.660	0.223	0.039	0.375	0.086	0.029	1.232	1.031	0.042	0.070	0.088	0.930	0.175	0.013	0.515	0.066	
5	WE dry paste	0.151	0.007	1.423	3.508	1.555	0.624	1.691	0.509	0.073	0.285	0.131	0.043	1.135	0.791	0.312	0.352	0.293	1.298	0.300	0.012	0.041	0.030	
6	70M dry paste	0.155	0.008	1.475	3.880	1.581	0.811	1.755	0.609	0.102	0.385	0.168	0.046	1.858	2.409	0.175	0.416	0.341	2.609	0.478	0.043	0.714	0.111	
7	70E dry paste	0.152	0.010	1.480	4.219	1.639	0.666	1.742	0.585	0.091	0.269	0.157	0.047	1.507	2.108	0.785	0.369	0.336	2.393	0.461	0.063	0.678	0.107	
8	XAD7HP resin	0.187	-0.018	4.423	17.298	6.519	3.514	4.450	2.530	0.420	2.439	0.884	0.326	10.159	11.048	4.086	1.562	1.345	12.231	2.775	0.205	3.515	0.469	
9																								
10																								
11																								
12																								
13																								
14																								
15																								
16																								
17																								
18																								
19																								
20																								

(a)

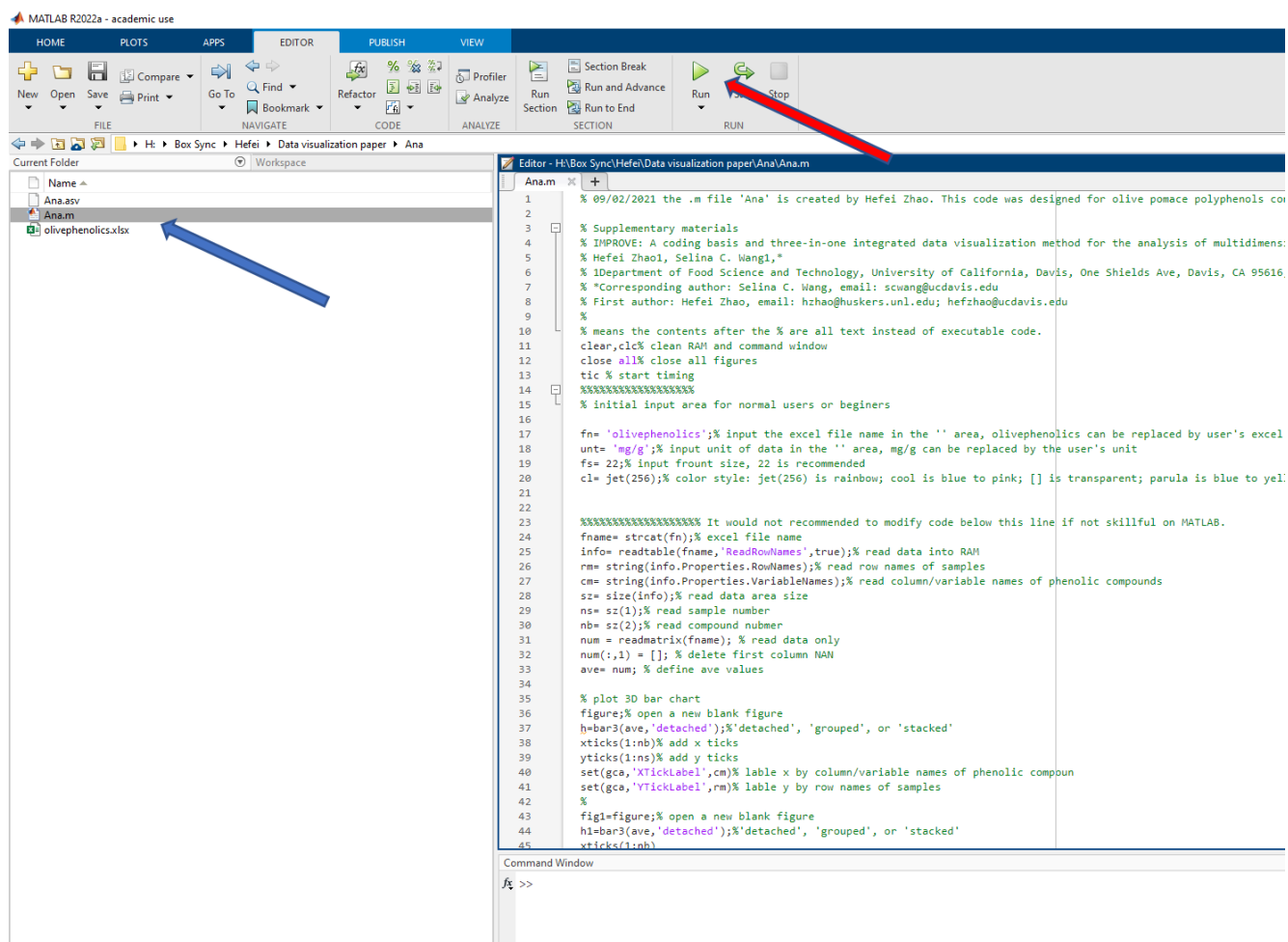


(b)

Figure S1. Data and code preparation. (a) Data from S. Table 1 was prepared in an excel file for reading from MATLAB, with one sample in each row, and one compound in each column. The 'NAME' in the cell of the first row and first column must be lined up at cell A1 in the excel file. Save as file 'olivephenolics.xlsx' in a folder. (b) Installation of data file and .m code from MATLAB 'Add-Ons': press 'HOME' tab at upper left, then click 'Add-Ons' tab, then search '3in1 omics data visualization and analytical method', then select the first result and install it.



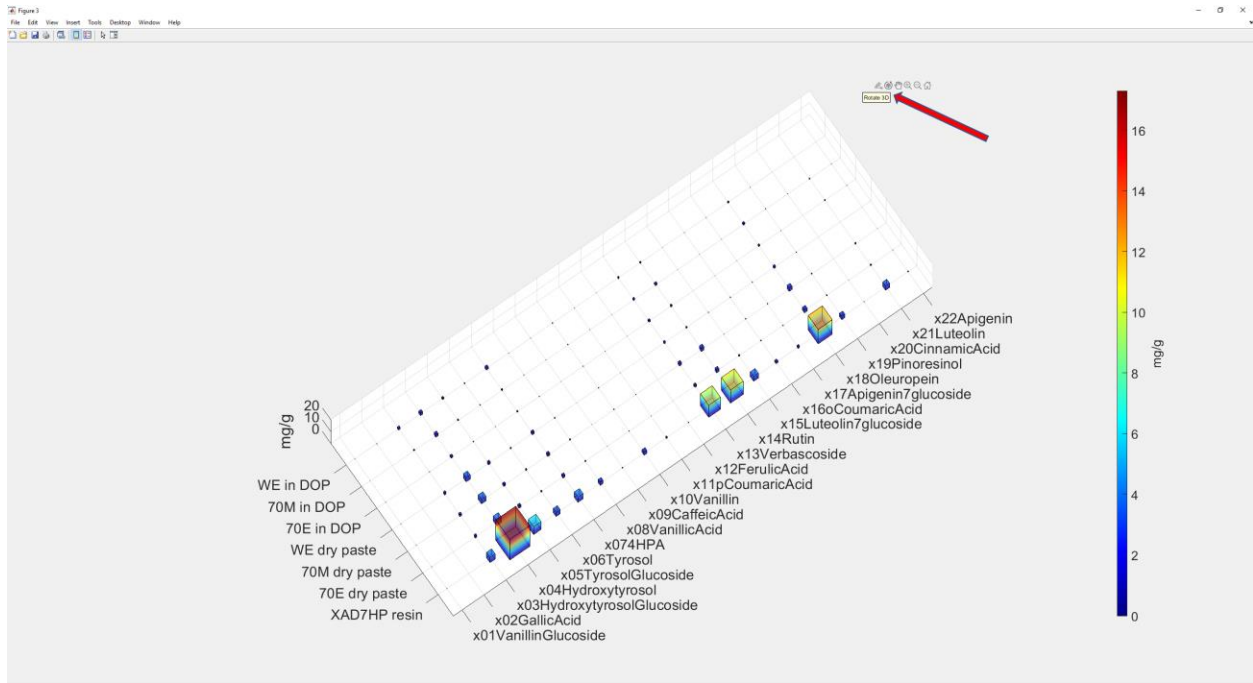
(a)



(b)

Figure S2. MATLAB .m file preparation. (a) click icon ‘+new’ (blue arrow) at the upper left of the software interface, then copy and paste the code **S. Code 1** into the editor window (red arrow), then save the Ana.m file in the same folder as the ‘olivephenolics.xlsx’ file; (b) both the excel ‘olivephenolics.xlsx’ file and the MATLAB ‘Ana.m’ file must be in the same folder; normally a ‘.asv’ file will be generated once click ‘run’ button to get results and figures

Note: Beside the Excel date and MATLAB ‘.m’ files can be reconstructed from the supplementary materials, the tool packages including all necessary files can also be downloaded from MATLAB file exchange website [1].



(a)

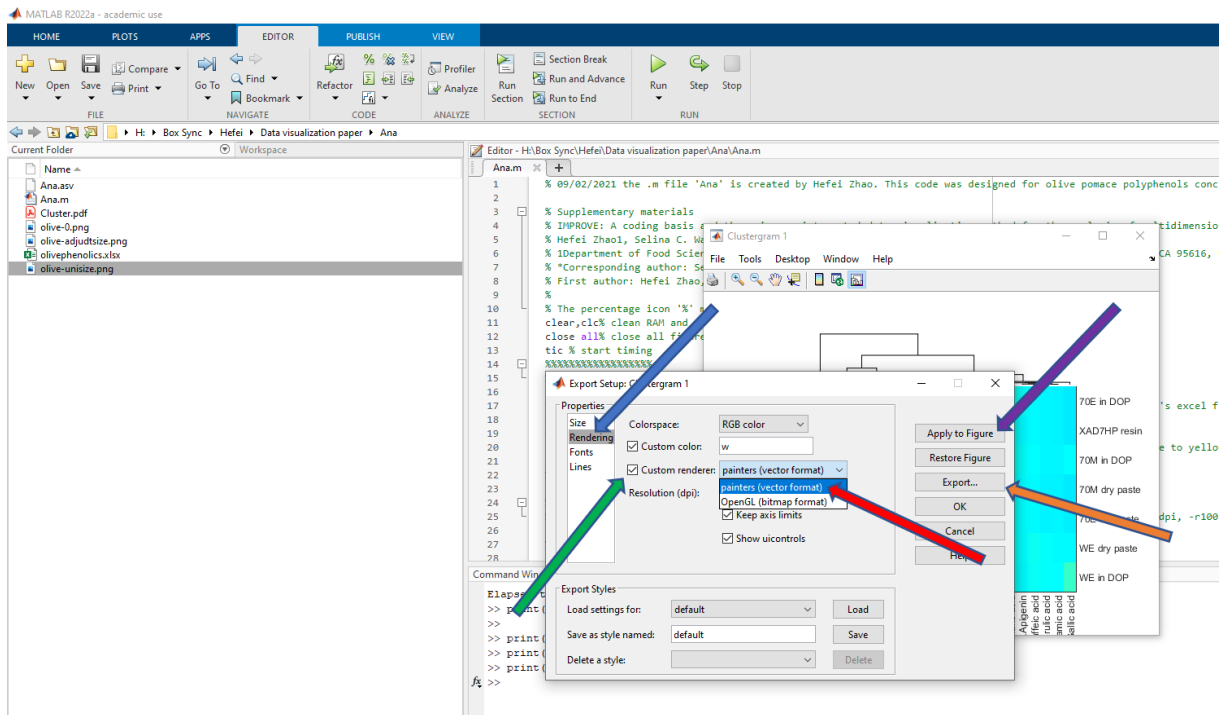
```

1 % 09/02/2021 the .m file 'Ana' is created by Hefei Zhao. This code was designed for olive p
2
3 % Supplementary materials
4 % IMPROVE: A coding basis and three-in-one integrated data visualization method for the ana
5 % Hefei Zhao, Selina C. Wangl,*
6 % Department of Food Science and Technology, University of California, Davis, One Shields
7 % *Corresponding author: Selina C. Wang, email: scwang@ucdavis.edu
8 % First author: Hefei Zhao, email: hzhao@huskers.unl.edu; hefzhao@ucdavis.edu
9
10 % The percentage icon '%' means the contents after the % are all text note instead of execu
11 clear,clc clean RAM and command window
12 close all close all figures
13 tic % start timing
14 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
15 % initial input area for normal users or beginners
16
17 fn= 'olivephenolics';% input the excel file name in the '' area, olivephenolics can be repl
18 unt= 'mg/g';% input unit of data in the '' area, mg/g can be replaced by the user's unit
19 fs= 22;% input front size, 18-22 are recommended
20 cl= jet(256);% color style: jet(256) is rainbow; cool is blue to pink; [] is transparent; p
21
22 % Print figures:
23
24 % First chart, 3D heatmap
25 % print(fig3Dbar,'olive-adjudsize.png','-dpng','-r150');% -r150 defines 150dpi,-r300 will
26
Command Window
Warning: Column headers from the file were modified to make them valid MATLAB identifiers be:
Set 'VariableNamingRule' to 'preserve' to use the original column headers as table variable :
Elapsed time is 3.753921 seconds.
>> print(fig3Dbar,'olive-adjudsize.png','-dpng','-r150')
>>
  
```

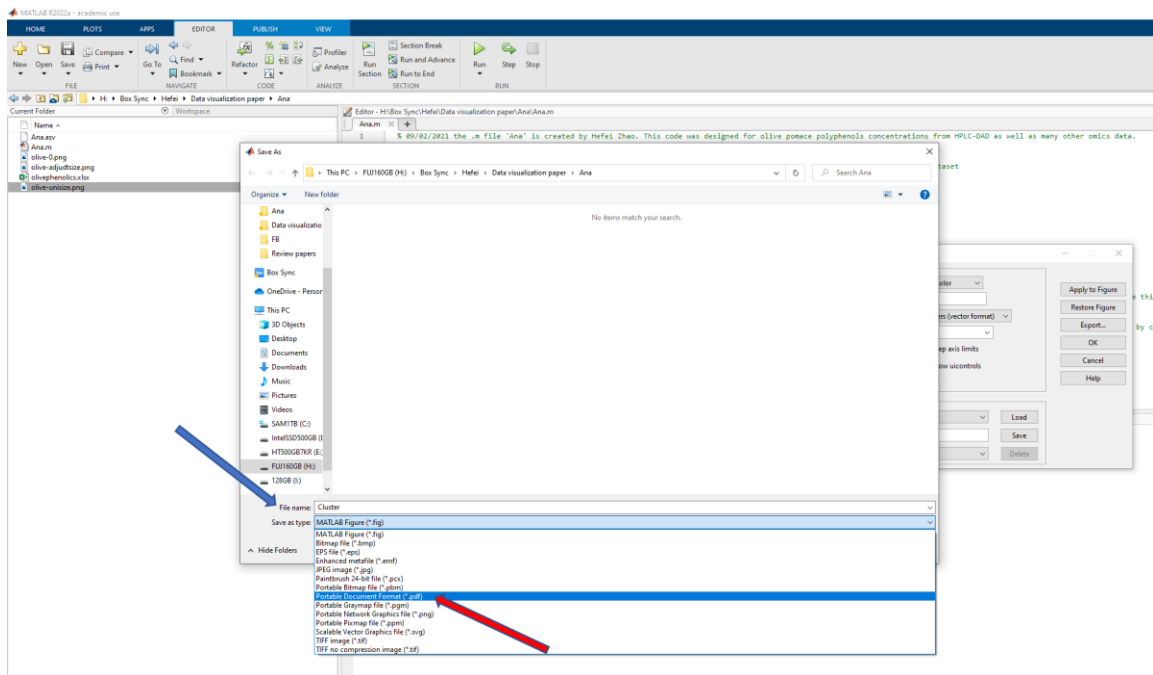
(b)

The screenshot displays the MATLAB R2022a environment. The top ribbon includes tabs for HOME, PLOTS, APPS, EDITOR, PUBLISH, and VIEW. The Editor window shows a script file named 'Ana.m' located at 'H:\Box Sync\Hefe\data visualization paper\Ana\Ana.m'. The script contains comments about its creation by Hefei Zhao and its purpose as supplementary materials for a coding basis. Below the script, a 'Clustergram' plot is visible, showing a heatmap with dendrograms on both axes. The File menu is open, and the 'Export Setup...' option is highlighted. A red arrow points from this option towards the Clustergram plot.

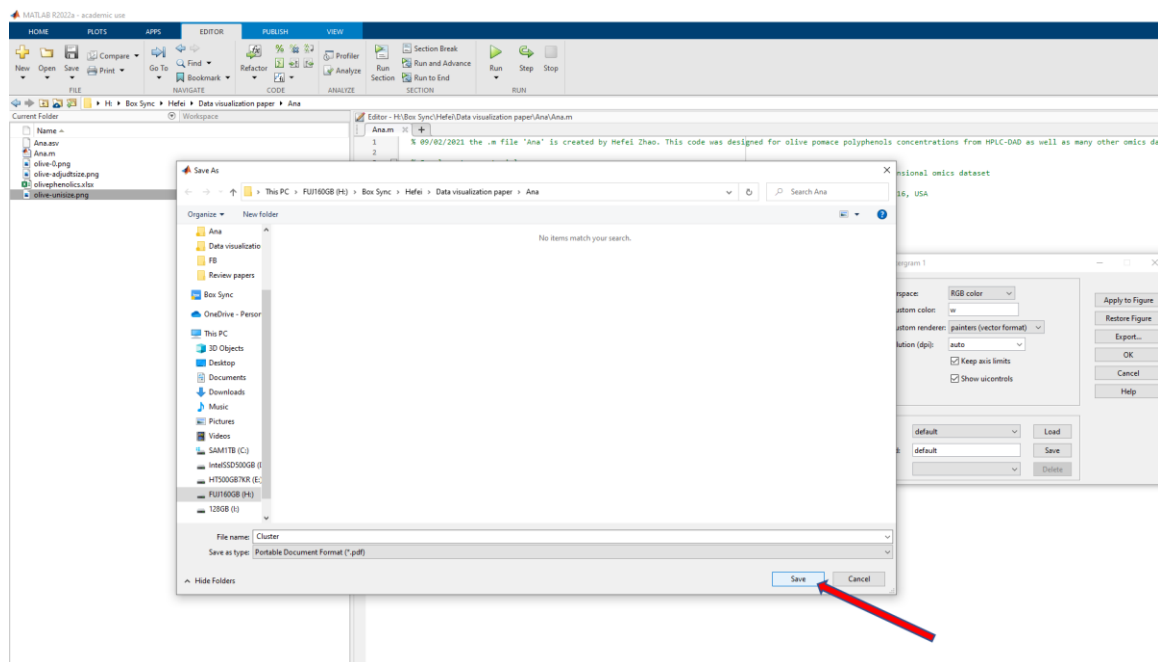
(b)



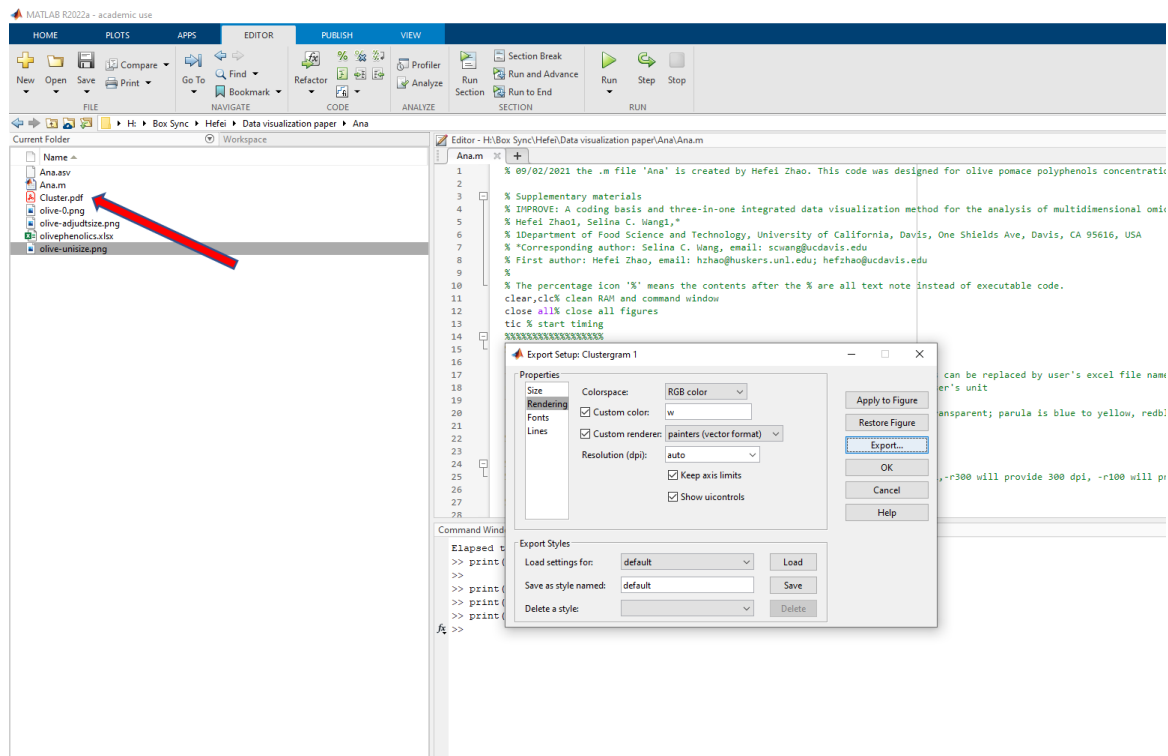
(c)



(d)



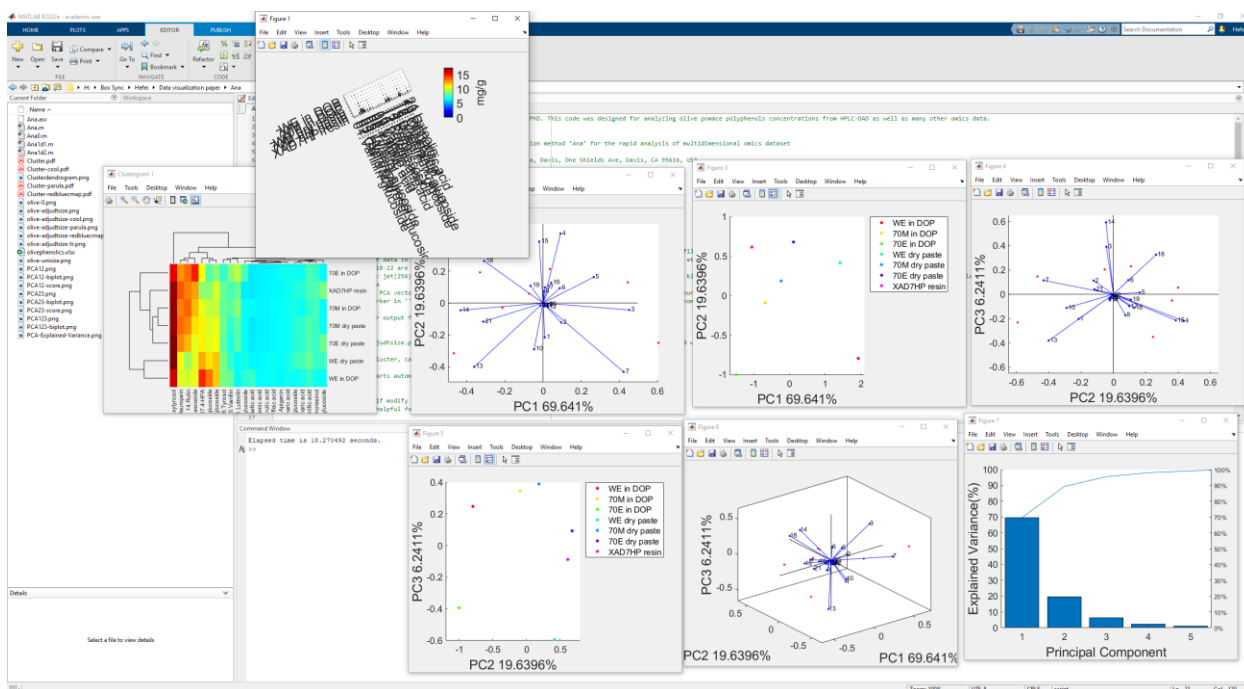
(e)



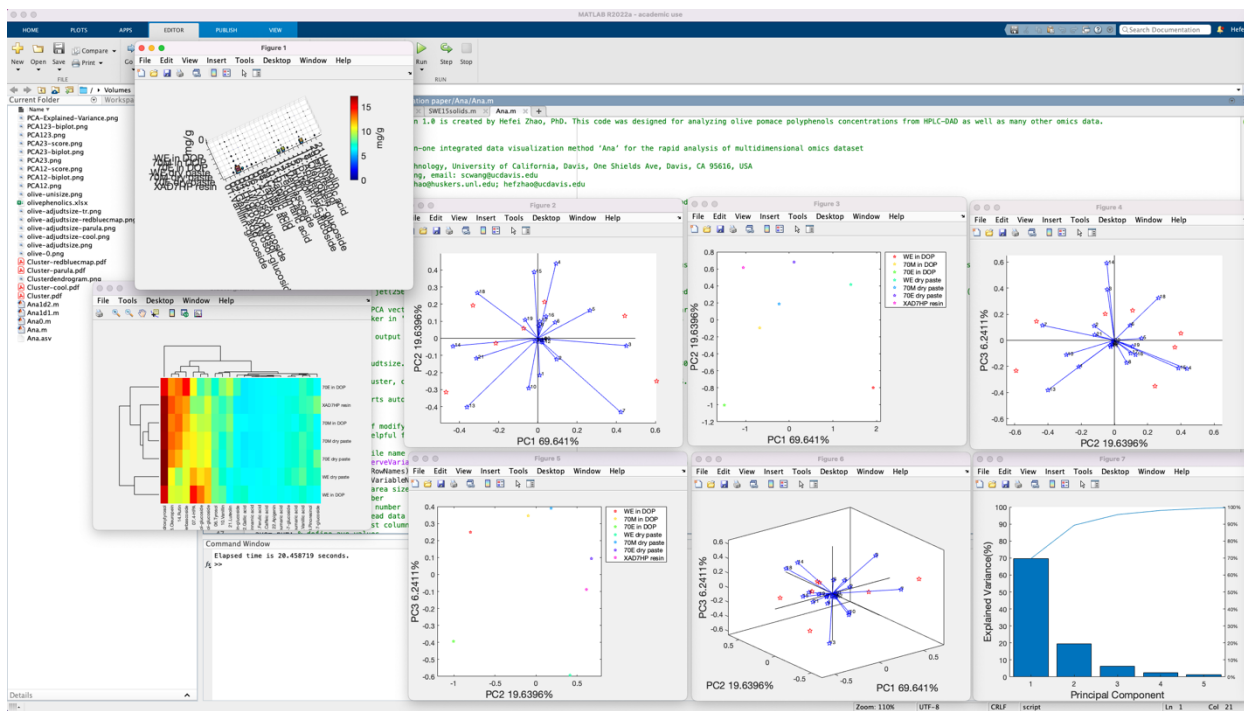
(f)

Figure S4. Export heatmap cluster chart to a '.pdf' file, (a) on the 'Clustergram 1' window, click 'Insert Colorbar', (b) on the 'Clustergram 1' window, click 'File', then click 'Export Setup', (c) click 'Rendering', select 'Custom rendering' as 'Painters (vector format)', very important for high-resolution output!!!, click 'Apply to Figure', then click 'Export', (d) input 'File name' as 'Cluster', select 'Save as type' the 'Portable Document Format (*.pdf)', (e)

click 'Save', (f) a 'Cluster.pdf' file will show up in the 'Current Folder'. Then open the '.pdf' file for 'print screen' a high-resolution figure.



(a)



(b)

Figure S5. Final outcomes of the program running. (a) data analyzed and output six figures in 18 seconds in the PC with Windows 10 system, (b) data analyzed and output six figures in 20 seconds in MacOS Monterey system.

S. Tables

Table S1. Phenolic compound data of olive pomace extract, data from our previous publication [2].

NAME	01.Vanillin-glucoside	02.Gallic acid	03.Hydroxytyrosol-glucoside	04.Hydroxytyrosol	05.Tyrosol-glucoside	06.Tyrosol	07.4-HPA	08.Vanillic acid	09.Caffeic acid	10.Vanillin	11.p-coumaric acid	12.Ferulic acid	13.Verbenascoside	14.Rutin	15.Luteolin-7-glucoside	16.o-coumaric acid	17.Apigenin-7-glucoside	18.Oleuropein	19.Pinoresinol	20.Cinnamic acid	21.Luteolin	22.Apigenin
WE in DOP	0.165	0.223	1.407	1.978	1.096	0.460	1.700	0.203	0.050	0.371	0.084	0.047	0.833	0.770	0.042	0.101	0.055	0.811	0.084	0.027	0.010	0.007
70M in DOP	0.054	0.045	0.657	2.017	0.679	0.365	0.800	0.208	0.044	0.329	0.097	0.023	1.074	1.360	0.042	0.070	0.121	1.270	0.257	0.019	0.487	0.062
70E in DOP	0.225	0.000	0.250	1.356	0.384	0.162	0.660	0.223	0.039	0.375	0.086	0.029	1.232	1.031	0.042	0.070	0.088	0.930	0.175	0.013	0.515	0.066
WE dry paste 70M	0.151	0.007	1.423	3.508	1.555	0.624	1.691	0.509	0.073	0.285	0.131	0.043	1.135	0.791	0.312	0.352	0.293	1.298	0.300	0.012	0.041	0.030
dry paste 70E	0.155	0.008	1.475	3.880	1.581	0.811	1.755	0.609	0.102	0.385	0.168	0.046	1.858	2.409	0.175	0.416	0.341	2.609	0.478	0.043	0.714	0.111
dry paste XAD7	0.152	0.010	1.480	4.219	1.639	0.666	1.742	0.585	0.091	0.269	0.157	0.047	1.507	2.108	0.785	0.369	0.336	2.393	0.461	0.063	0.678	0.107
HP resin	0.187	-0.018	4.423	17.298	6.519	3.514	4.450	2.530	0.420	2.439	0.884	0.326	10.159	11.048	4.086	1.562	1.345	12.231	2.775	0.205	3.515	0.469

Note: WE, water extract; 70M, 70% methanol extract; 70E, 70% ethanol extract; XAD7HP resin, XAD7HP resin purified extract

S. References

- 1 Zhao, H.; Wang, C.S. A 3in1 Omics Data Visualization and Analytical Method—File Exchange—MATLAB Central. Available online: <https://www.mathworks.com/matlabcentral/fileexchange/117370-a-3in1-omics-data-visualization-and-analytical-method> (accessed on 8 September 2022).
- 2 Zhao, H.; Avena-Bustillos, R.J.; Wang, S.C. Extraction, Purification and In Vitro Antioxidant Activity Evaluation of Phenolic Compounds in California Olive Pomace. *Foods* 2022, 11, 174.