Supplementary material

Table S1. List of pre-processing methods for hyperspectral data of *B. juncea* plants.

1st pre-processing	2nd pre-processing	3rd pre-processing	Abbreviation
Raw reflectance	None	None	Raw reflectance
		SNV	SNV
		MSC	MSC
	Savitzky–Golay filter	None	SG filter
		SNV	SG filter + SNV
		MSC	SG filter + MSC
	1st derivative	None	1st Der
		SNV	1st Der + SNV
		MSC	1st Der + MSC
	2nd derivative	None	2nd Der
		SNV	2nd Der + SNV
		MSC	2nd Der + MSC
Normalization	None	None	Norm
		SNV	Norm + SNV
		MSC	Norm + MSC
	Savitzky–Golay filter	None	Norm + SG filter
		SNV	Norm + SG filter + SNV
		MSC	Norm + SG filter + MSC
	1st derivative	None	Norm + 1st Der
		SNV	Norm + 1st Der + SNV
		MSC	Norm + 1st Der + MSC
	2nd derivative	None	Norm + 2nd Der
		SNV	Norm + 2nd Der + SNV
		MSC	Norm + 2nd Der + MSC
Logarithmic transformation	None	None	Log (1/R)
		SNV	Log (1/R) + SNV
		MSC	Log (1/R) + MSC
	Savitzky–Golay filter	None	Log (1/R) + SG filter
		SNV	Log (1/R) + SG filter + SNV
		MSC	Log (1/R) + SG filter + MSC
	1st derivative	None	Log (1/R) + 1st Der
		SNV	Log (1/R) + 1st Der + SNV
		MSC	Log (1/R) + 1st Der + MSC
	2nd derivative	None	Log (1/R) + 2nd Der
		SNV	Log (1/R) + 2nd Der + SNV
		MSC	Log (1/R) + 2nd Der + MSC

SNV, standard normal variate; MSC, multiplicative scatter correction

Table S2. Determination of pre-processing methods for Adaboost, XGboost, and LightGBM prediction algorithms for five metabolites in *B. juncea* plants.

Metabolite	Prediction algorithms	Pre-processing method	Calibration		Cross-validation	
			R^2c	RMSEC	R^2cv	RMSECV
Chlorophylls	**Adaboost**	**Log (1/R) + 2nd Der + MSC**	**0.934**	**0.593**	**0.425**	**1.750**
		1st Der	0.921	0.648	0.420	1.757
		Log (1/R) + 1st Der	0.933	0.595	0.404	1.782
		Log (1/R) + 2nd Der	0.929	0.614	0.388	1.805
		Log (1/R) + 2nd Der + SNV	0.934	0.594	0.386	1.808
	XGboost	**Log (1/R) + 2nd Der + MSC**	**1.000**	**0.000**	**0.449**	**1.713**
		1st Der	1.000	0.000	0.430	1.742
		Log (1/R) + 2nd Der + SNV	1.000	0.000	0.412	1.770
		Log (1/R) + 2nd Der	1.000	0.000	0.401	1.785
		Norm + MSC	1.000	0.001	0.391	1.800
	LightGBM	**1st Der**	**0.945**	**0.539**	**0.402**	**1.783**
		Log (1/R)	0.830	0.950	0.382	1.814
		Log (1/R) + 1st Der + SNV	0.942	0.553	0.375	1.823
		Log (1/R) + 2nd Der	0.969	0.403	0.361	1.844
		Log (1/R) + 2nd Der + SNV	0.967	0.422	0.346	1.866
Phenolics	**Adaboost**	**Norm**	**0.924**	**0.642**	**0.581**	**1.505**
		Norm + MSC	0.913	0.687	0.540	1.576
		Log (1/R) + 1st Der + MSC	0.937	0.581	0.537	1.582
		1st Der + MSC	0.934	0.598	0.536	1.584
		Norm + 1st Der	0.937	0.582	0.528	1.597
	XGboost	**Norm + SG filter**	**1.000**	**0.001**	**0.574**	**1.516**
		Log (1/R) + 1st Der + MSC	1.000	0.000	0.560	1.541
		SNV	1.000	0.001	0.555	1.551
		Norm + SNV	1.000	0.001	0.555	1.551
		Norm + MSC	1.000	0.000	0.552	1.555
	LightGBM	**1st Der**	**0.951**	**0.515**	**0.522**	**1.607**
		Log (1/R) + SG filter + MSC	0.902	0.726	0.502	1.639
		Norm + 2nd Der	0.974	0.378	0.502	1.639
		SG filter + SNV	0.916	0.675	0.499	1.645
		Norm + SG filter + SNV	0.916	0.675	0.499	1.645
Flavonoids	**Adaboost**	**2nd Der**	**0.924**	**0.676**	**0.467**	**1.789**
		Raw reflectance	0.842	0.974	0.446	1.824
		Log (1/R)	0.838	0.986	0.441	1.832
		1st Der	0.927	0.663	0.435	1.842
		Log (1/R) + 1st Der + SNV	0.919	0.696	0.433	1.846
	XGboost	**1st Der**	**1.000**	**0.000**	**0.521**	**1.696**
		Norm + SG filter + MSC	1.000	0.001	0.508	1.719
		2nd Der	1.000	0.000	0.493	1.744
		SG filter + SNV	1.000	0.001	0.464	1.793
		Norm + SG filter + SNV	1.000	0.001	0.464	1.793
	LightGBM	**1st Der**	**0.930**	**0.650**	**0.480**	**1.766**
		MSC	0.910	0.735	0.445	1.825
		Norm + SNV	0.909	0.740	0.443	1.829
		SNV	0.909	0.740	0.443	1.829
		Norm + MSC	0.907	0.748	0.439	1.835
Glucosinolates	**Adaboost**	**SNV**	**0.935**	**2.868**	**0.666**	**6.481**
		Norm + SNV	0.935	2.868	0.666	6.481
		SG filter + SNV	0.929	2.996	0.657	6.572

Metabolite	Algorithm	Pre-processing method				
		Norm + SG filter + SNV	0.929	2.996	0.657	6.572
		SG filter	0.913	3.317	0.653	6.605
	XGboost	**SG filter + SNV**	**1.000**	**0.001**	**0.667**	**6.473**
		Norm + SG filter + SNV	1.000	0.001	0.667	6.473
		SNV	1.000	0.001	0.660	6.543
		Norm + SNV	1.000	0.001	0.660	6.543
		1st Der	1.000	0.000	0.649	6.648
	LightGBM	**Log (1/R) + 1st Der + SNV**	**0.972**	**1.883**	**0.672**	**6.420**
		1st Der	0.964	2.114	0.663	6.512
		Log (1/R) + 1st Der + MSC	0.974	1.805	0.659	6.550
		2nd Der	0.983	1.443	0.658	6.561
		SG filter	0.910	3.363	0.657	6.568
Anthocyanins	**Adaboost**	**Log (1/R) + 1st Der + MSC**	**0.972**	**0.930**	**0.822**	**2.330**
		Log (1/R) + SG filter + SNV	0.971	0.932	0.734	2.843
		Norm + 1st Der + SNV	0.976	0.850	0.708	2.980
		1st Der + SNV	0.976	0.850	0.708	2.980
		Norm + 1st Der	0.973	0.903	0.704	3.003
	XGboost	**1st Der**	**1.000**	**0.000**	**0.720**	**2.920**
		Norm + 1st Der + SNV	1.000	0.000	0.701	3.014
		1st Der + SNV	1.000	0.000	0.701	3.014
		2nd Der	1.000	0.000	0.665	3.194
		1st Der + MSC	1.000	0.000	0.643	3.296
	LightGBM	**Log (1/R) + 1st Der**	**0.901**	**1.733**	**0.688**	**3.081**
		1st Der + SNV	0.900	1.746	0.679	3.125
		Norm + 1st Der + SNV	0.900	1.746	0.679	3.125
		Norm + 1st Der	0.903	1.717	0.673	3.153
		Norm + 1st Der + MSC	0.904	1.713	0.661	3.214

Only the top five pre-processing methods out of 36 methods for each metabolite and prediction algorithm are shown. Bold indicates the pre-processing methods determined with the lowest RMSECV.