Journal of Agricultural and Food Chemistry, 2017, Volume 65, Issue 26, pp5339-5356
DOI:10.1021/acs.jafc.7b00500

Biochemical Profile of Heritage and Modern Apple Cultivars and Application of Machine Learning Methods to Predict Usage, Age, and Harvest Season

Maria Anastasiadi[1], Fady Mohareb[1], Sally P. Redfern[2], Mark Berry[2], Monique S.J. Simmonds[3], Leon A. Terry*[1]

[1]Plant Science Laboratory, Cranfield University, Bedfordshire, UK, MK43 0AL

[2]Unilever R&D Colworth, Sharnbrook, Bedfordshire, UK, MK44 1LQ

[3]Royal Botanic Gardens, Kew, Ricmond, Surrey, UK, TW9 3AB

Published by American Chemical Society. This is the Author Accepted Manuscript issued with:
Creative Commons Attribution Non-Commercial License (CC:BY:NC 4.0).
The final published version (version of record) is available online at DOI:10.1021/acs.jafc.7b00500.
Please refer to any applicable publisher terms of use.

ABSTRACT

The present study represents the first major attempt to characterise the biochemical profile in different tissues of a large selection of apple cultivars sourced from the UK's National Fruit Collection comprising dessert, ornamental, cider and culinary apples. Furthermore, advanced Machine Learning methods were applied with the objective to identify whether the phenolic and sugar composition of an apple cultivar could be used as a biomarker fingerprint to differentiate between heritage and mainstream commercial cultivars as well as govern the separation among primary usage groups and harvest season. Prediction accuracy > 90% was achieved with Random Forest for all three models. The results highlighted the extraordinary phytochemical potency and unique profile of some heritage, cider and ornamental apple cultivars, especially in comparison to more mainstream apple cultivars. Therefore, these findings could guide future cultivar selection on the basis of health-promoting phytochemical content.

KEYWORDS: *Malus*, phenolic compounds, sugars, organic acids, amygdalin, predictive modelling

INTRODUCTION

Malus domestica is one of the most widely cultivated tree fruits with great economic and cultural value,[1] and the most widely known of the many members of the genus *Malus.* At present, a significant amount of literature exists on the nutritional and phytochemical content of apples both for quality assessment and determining the levels of compounds with potential health promoting properties such as ascorbic acid and phenolic compounds.[2–5] Most studies have focused on commercial dessert apple cultivars where the phytochemical profile is relatively well established. In contrast, only a few references exist on the biochemical content of underutilised heritage cultivars and apple cultivars not intended for fresh consumption, such as culinary and cider apples. Even less information is available on other species of *Malus* generally known as crab-apples or wild apples which are closely related to the domesticated apple. When the apple genome was published in 2010, the wild Central Asian species *Malus sieversii* was identified as the main contributor to the genome of the cultivated apple.[6] However, other recent DNA analysis has revealed that multiple species have contributed to the genetic makeup of domesticated apples, with the wild European crab-apple *Malus sylvestris* in particular, being a major secondary contributor.[7] Crab-apple species are popular as ornamental plants generally bearing small to medium size fruits with a characteristic bitter and astringent taste. Some crab-apples can be used for culinary purposes, mainly due to the often vibrant red colour of their flesh and peel, which is the result of high expression levels of specific genes related to anthocyanin accumulation.[8] Cider apples have been proposed to be directly linked to *M. sylvestris* crab-apples as they may have been specifically selected, for their particular organoleptic properties during domestication, for the preparation of beverages such as cider.[1] Cider making with crab-apples was known in Western Europe, before the introduction of the domesticated apple by the Romans. It was only in the 17th century that efforts were

intensified to breed cultivars high in phenolic compounds and sugars for the production of high quality cider.[9]

Although more than 7,500 varieties of apples exist worldwide, many heritage varieties have been abandoned, despite a resurgence observed in recent years, in favour of mainstream varieties emerging from intensive selective breeding programmes during the last few decades leading to the decline of traditional apple orchards in many countries including the UK. The reason for this is the drive for sweet, crisp apples, which are uniform in size and appearance, and have acceptable disease resistance and prolonged shelf-life.

In the United Kingdom, some older cultivars such as 'Cox's Orange Pippin', 'Egremont Russet' and 'Bramley' are still commercially important but most other heritage cultivars have experienced a rapid decline. As a result many ancient cultivars have been irreversibly lost. A number of initiatives have arisen recently for the preservation of plant genetic resources for the future, such as the Millennium Seed Bank Partnership and the 'International Treaty on Plant Genetic Resources for Food and Agriculture' (2004). The National Fruit Collection (NFC) in the U.K. is part of this international programme to protect plant genetic resources for the future, and hosts more than 2700 accessions of apples many of which are heritage cultivars dating back to the 13th century. Apart from dessert apple cultivars the NFC also includes a wide selection of culinary, cider and ornamental (crab) apples.

Selective breeding programmes are influenced by consumer and industry demand for the quality traits above. The resulting lines could therefore have different levels of certain compounds such as sugars, phenolic compounds and organic acids which give each apple cultivar their characteristic taste. Phenolic compounds in particular are widespread secondary metabolites which contribute to the colour and taste characteristics of apples, such as bitterness and astringency; traits which are not always desirable in a modern dessert

apple. On the other hand, phenolic compounds, sugars and organic acids are important for the development of the characteristic taste of cider. Thus, cider apples have been traditionally selected for these traits and have been generally categorised as 'sharps', 'sweets', 'bittersweets' and 'bittersharps', depending on the sugar/acid/tannin ratio.[9]

Apart from their role in taste and appearance, apple phenolic compounds have been associated with health promoting properties,[5,10,11] with apple being one of the major sources of dietary polyphenols worldwide. A complex range of phenolic compounds is present in apples, including hydroxycinnamic acids, flavan-3-ols and oligomeric procyanidins, dihydrochalcones, flavonols and anthocyanins.[3,4,12] Phloridzin in particular, a glycoside of phloretin which is a characteristic compound of *Malus* species,[13] has attracted a lot of attention due to its potential antidiabetic properties, as it is a known sodium-dependent glucose co-transporter-1 (SGLT1) and sodium-dependent glucose co-transporter-2 (SGLT2) inhibitor.[11,14]

Literature reports suggest that heritage apple cultivars have remarkably different phenolic profiles compared to mainstream cultivars with the former tending to have increased concentrations of certain phenolic subclasses; mainly flavanols and procyanidins, dihydrochalcones and hydroxycinammic acid derivatives.[2,12,15] The present study sought to answer whether this phenomenon is a universal trend by examining a wide selection of underutilised heritage and mainstream commercial cultivars and applying advanced machine learning techniques in an attempt to distinguish between heritage and modern cultivars based on their phenolic and sugar profile. In addition, the same techniques were further applied to distinguish between apple cultivars based on usage and harvest season. Machine learning approaches have found an increasing number of applications in food science and agriculture in recent years with examples including chemometric spectral analysis and targeted metabolomics simulations. For instance, ensemble-based Support

Vector Machines (SVM) in tandem with electronic nose has been applied to develop freshness prediction models for meat products.[16] Artificial Neural Networks (ANN) have successfully been used for the prediction of food quality and have been shown to perform equally well to statistically based prediction methods such as Partial Least Squares (PLS).[17] Other methods such as Random Forests (RF) have found applications in the prediction of crop yield[18] and future crop cover patterns associated with climate change.[19] In this context, the current study represents an attempt to capture the biochemical diversity within the apple breeding pool and to identify cultivars with distinctive qualities which could guide future breeding programmes for cultivars with enhanced health promoting properties.

MATERIALS AND METHODS

Plant Materials and Sample Preparation. A total of 66 apple cultivars were collected at commercial maturity over three years (2012-2014) from Kent, United Kingdom. Cultivars were harvested from the National Fruit Collection, Brogdale Farm, Kent, UK (51°, 18′S, 0°, 52′E) and Worldwide Fruit Ltd, Kent, (51°, 21′S, 1°, 3′E). Maturity was assesed by experienced farm staff based on fruit appearance (size, colour of the skin, flesh, seeds and flavour where applicable) and information related to picking times for each cultivar contained in the NFC database. All apple trees in the NFC were grafted on M.9 rootstocks (dwarf trees, 2-2.5 m high) and planted on 2-tree plots per cultivar and grown under semi-commercial standards. The selected material comprised heritage and modern dessert apple cultivars as well as culinary, cider and ornamental cultivars. The harvest period spanned between August to November covering apple cultivars maturing at different times and under different weather conditions. Many of the cultivars considered in this study, represent very old apple cultivars and to our knowledge their biochemical profile has never been reported. ‘Decio’ is perhaps the oldest cultivar included in the NFC, and is believed to have been

brought to England by the Romans. 'Old Pearmain (of Kelsey)' is another example of a very old cultivar, first recorded around 1200 in both UK and France and believed to have been brought to England by the Normans. The 11 apple cultivars supplied in 2012 from Worldwide Fruit Ltd included 'Bramley', 'Worcester Pearmain', 'Early Windsor', 'Queen Cox', 'Royal Gala', 'Spartan', 'Falstaff', 'Ashmeads Kernel', 'Jazz', 'Braeburn' and 'Golden Delicious'. Collection of some of the above cultivars was repeated in 2013 and 2014 exclusively from the NFC. Further information is included in **Table 1**. Twenty-four apples were randomly picked from both the paired cultivars in the NFC or from the commercial orchards. After harvest the fruits were delivered to the lab within 24 h where they were stored at 5 °C and processed within two days. Twelve fruits per cultivar were selected and assessed for the following parameters: a) height, and diameter; b) objective colour (L*,C*,H°), using a hand-held Minolta colorimeter (CR-400 Chroma Meter, Konica Minolta Inc, Warrington, UK). For each apple the objective colour was separately assessed for the whole fruit, the light exposed and shaded side for non-pigmented fruits and the red and green side for pigmented fruits. The results are presented in the supplementary material (**Supporting material S1**).

Each apple was sectioned as follows: an equatorial slice (approx. 10 mm thickness), was cut from each fruit and the seeds were removed and snap-frozen with liquid N_2. The equatorial slice, representative of the whole fruit (edible part), was diced and immersed in liquid N_2. The remaining top and bottom parts of each apple were divided into peel and flesh and each tissue was snap-frozen separately with liquid N_2. The above procedure was performed as quickly as possible to avoid any browning occurring. All samples were stored at –80 °C until further analysis.

The remaining fruits were further assessed for firmness and maturity. Firmness was measured with a uniaxial testing machine (Instron 5542, Instron, Buckinghamshire, UK). A 10 mm

diameter probe was used at 240 mm/min cross head speed and 8 mm penetration depth. Firmness was assessed at two opposite positions for each apple, the light exposed side and the shaded side (**Supporting material S1**). Maturity was assessed using the starch index and rating hydrolysis of starch on a scale from 1 (100% starch) to 10 (0% starch) for 6 apples per cultivar (**Supporting material S2**).

Chemicals All HPLC and LC-MS grade solvents were obtained from Fisher Scientific (Loughborough UK). (+)-Catechin, (–)-epicatechin, procyanidin B1, procyanidin B2, chlorogenic acid, cryptochlorogenic acid, caffeic acid, *p*-coumaric acid, quercetin-rutinoside (rutin), quercetin-glucoside, quercetin-galactoside, phloridzin dihydrate were purchased from Sigma-Aldrich (Dorset, UK). Cyanidin-3-O-galactoside (ideain) chloride, quercetin-3-xyloside, quercetin-3-*O*-α-L-arabinofuranoside (avicularin) were purchased from Extrasynthese (Genay Cedex, France). Metaphosphoric acid (Bioxtra ≥ 33.5%), phosphoric acid (BioUltra, ≥ 85%), potassium phosphate monobasic, D-fructose, D-sorbitol, L(-)-ascorbic acid, oxalic acid, tartaric acid, quinic acid, malic acid, maleic acid, shikimic acid, citric acid, succinic acid and fumaric acid were obtained from Sigma-Aldrich (Dorset, UK). D-glucose and sucrose, were purchased from Fisher Scientific (Loughborough UK). Iodine and potassium iodide (BioUltra, ≥ 99.5%) were obtained from Sigma-Aldrich (Dorset, UK).

Phenolic compounds. Phenolic compounds for the whole fruit were analysed for all samples over three years. In addition, the phenolic profile of each separate tissue (peel, flesh, seeds) was analysed for 20 cultivars from 2012. The rationale behind this selection was to aquire a representative sub-sample of all the main groups studied (usage, age) and examine how each tissue influences the total content and whether the striking differences in concentrations found in the whole apple were also reflected in each separate tissue.

Before extraction, samples were freeze-dried and each tissue was separately powdered in a mortar grinder (RM 200, Retsch Ltd., Derbyshire, UK). Tissue powder from six individual

apples per cultivar (biological replicates) was separately extracted with the following protocol: freeze-dried whole apple tissue (300 mg), flesh tissue (150 mg), peel tissue (50 mg) and seed tissue (30 mg) were extracted for 15 min with 6 mL, 3 mL, 1 mL and 0.6 mL respectively of 70% (v/v) aqueous acetone (0.1% formic acid) in a water bath at 35 ºC with frequent mixing to re-suspend the solids. The samples were centrifuged at 4000 rpm and the supernatant was removed. The extraction process was repeated twice and the organic layers were combined. The organic solvent was subsequently evaporated using a centrifugal vacuum concentrator (miVac Quatro, Genevac Ltd, Suffolk, UK) and the remaining aqueous phase was further extracted with 2 x 4 mL hexane. The organic layer was removed and the aqueous phase was freeze-dried to dryness. The final extract was reconstituted with 2 mL (or 1 mL for the seeds) of 70% aqueous methanol (0.1% formic acid) and filtered through a 0.2 μm PTFE filter. The extracts were diluted further with mobile phase just before analysis. Characterisation and identification of phenolic compounds was based on comparison with commercial standards and by obtaining their accurate mass profile on an Agilent Ultra High Definition Accurate Mass Q-TOF–MS system (Agilent Technologies LDA Cheshire, UK) equipped with an electrospray ionization source (Agilent Dual Jet Stream) and coupled with an Agilent 1290 infinity UPLC system, comprised of a binary pump with a jet weaver V35 mixer, a thermostated column, set at 30 °C, a cooled autosampler set at 6 °C with a previously reported method[20]. Identification was further aided by comparing the RT and UV/Vis spectra of apple extract phenolic compounds with that of commercial standards on an Agilent 1200 HPLC with a DAD detector fitted with an Eclipse XDB-C18 column (150 mm x 4.6 mm; 5 μm, Agilent Technologies) and a 1mm OPTI-Guard Column. The mobile phase consisted of solvent A (5% formic acid in HPLC water) and solvent B (acetonitrile) and the elution gradient was as follows: 0-30 min, 0-10% B, 30-48 min, 10-30% B, 48-50 min, 30-100% B; 50-55 min, 100% B, followed by 5 min re-equilibration time. Detection was performed in 4 different wavelengths, 280 nm, 320 nm, 360 nm and 520 nm.

Phloretin-2'-*O*-xylosylglucoside detected at 280 nm, exhibited identical UV/Vis profile with phloridzin and phloretin. Its mass spectrum had a [M-H]+ profile with *m/z* 569.1873 (58.14% relative abundance), corresponding to the molecular ion of phloretin-2'-*O*-xylosylglucoside, *m/z* 591.1695 (68.73% relative abundance), corresponding to the [M-Na]+ adduct, and *m/z* 275.0919 (100% relative abundance) corresponding to the phloretin aglycon molecular ion. *p*-Coumaroyl quinic acid detected at 320 nm had UV/Vis spectra similar to *p*-coumaric acid and its mass spectrum had a [M-H]+ profile with *m/z* 339.1074 (58.14% relative abundance), corresponding to the molecular ion of *p*-coumaroyl quinic acid, *m/z* 361.0892 (30.79% relative abundance) corresponding to the [M-Na]+ adduct and also 165.0543 (7.71% relative abundance) corresponding to the molecular ion of *p*-coumaric acid and *m/z* 147.0439 (100% relative abundancy) a characteristic fragment of *p*-coumaric acid.

Quantification was performed on the HPLC-DAD system and was based on external calibration curves of commercial standards. These standards included the most abundant phenolic compounds belonging to each of the main classes of phenolic compounds present in apples.

UPLC/QTOF/MS analysis of seed extracts. The amygdalin and phenolic content of the seed extracts was analysed using the UPLC-QTOF-MS method described above. Mass spectra were recorded in negative ion mode between 100 and 1500 atomic mass units (amu), except for anthocyanins which were analysed in positive ion mode.

Soluble sugars. Extraction and analysis of soluble sugars was performed for all the apple samples collected over 3 years, using a previously described method[20] with slight modifications. Briefly, 150 mg of whole apple freeze-dried powder were extracted for 15 min with 3 mL of 62.5% (v/v) aqueous methanol (in a water bath at 55 ^{o}C with frequent mixing to re-suspend the solids. Prior to analysis, the sugars extracts were diluted (1:9 v/v) with HPLC water. The eluted compounds were detected by Evaporative Light Scattering Detector (ELSD) and quantification was based on external calibration curves of commercial standards.

Non-volatile organic acids. Non-volatile organic acids were analysed for the sub-group of 20 cultivars from 2012, assessed for the spatial distribution of phenolic compounds. The method used was a modification of a previously described method[20] for the analysis of total ascorbic acid. Briefly, fresh whole apple tissue stored at -80 °C, was powdered in a mortar grinder (RM 200, Retsch Ltd., Derbyshire, UK) with liquid N_2. Next, 1g frozen tissue was extracted with 5 mL metaphosphoric acid (0.01 M) in a shaking water bath at 25 °C for 10 min. The resulting slurry was filtered through cellulose acetate filters 0.2 μm and immediately injected in an Agilent 1200 HPLC with a DAD detector fitted with a GRACE Altima HP C18 AQ column (150 mm x 4.6 mm; 5 μm, GRACE). The mobile phase consisted of potasium phosphate monobasic solution (0.25 mM) adjusted to pH 2.5 with phosphoric acid and the elution time was set to 10 min. All organic acids were monitored at 210 nm except ascorbic acid which was monitored at 248 nm. Quantification of organic acids was based on external calibration curves of commercial standards.

Statistical Analysis. Analysis of variance was performed using Genstat for Windows, Version 12 (VSN International Ltd., Herts., UK) and hierarchical cluster analysis (HCA) with the Fast Ward method was performed using JMP 13 (SAS Institute Inc. Bucks, UK). The differences between means of data were compared through Least Significant Difference (lsd) and they were considered to be statistically significant at the 95% confidence level ($p \leq 0.05$). Concentrations below the limit of quantification (LOQ) were replaced by ½ of LOQ for the respective compound. Logarithmic transformations were employed where needed in order to ensure the assumption of equal variability.

Classification Modelling. A number of pattern recognition and machine learning techniques, namely k-Nearest Neighbours (*k*NN), Naïve Bayes (NB), Support Vector Machines (SVM) and Random Forest (RF), were applied to develop predictive models to classify apple cultivars

according to age, usage and harvest season, based on the biochemical profile of the whole apple extract.

*k*NN, is a machine learning technique which applies sample distance to perform classification.[21] Briefly, the k-closest points to the sample are considered, before a majority vote is applied to classify it or predict its value. *k*NN was implemented using the function "knn" from the "FNN" R package.[22] The best *k* was selected using a grid search from k=4 to 10. NB is a probabilistic method calculating the probability of an event occurring given the probability of another event that has already occured; known as conditional probability. The algorithm is based on the posterior probability of the sample belonging to each of the classes by combining (multiplying) the prior probability of belonging to one class by the likelihood of the new sample belonging to such class. RF is an ensemble method based on bootstrap aggregation.[23] This method constructs multiple versions of the training data by sampling with replacement (bootstrapping), creates a model and makes predictions for all of them and combines the predictions. The RF algorithm uses bootstrap samples, creates tree models for a certain number of random features for each one of the bootstrap samples and predictions of the tree models are combined to obtain the final prediction. RF was implemented with 200 trees using the "randomForest" function from the "randomForest" R package.[24] SVM is a supervised learning method for object classification in n-dimensional hyperspace while advances in optimisation and generalization methods are used to increase efficiency and prevent 'over-fitting'. To find the best values for these parameters a grid search was carried out. The implementation was performed using the "svm" function from the "e1071" R package.[25]

For the age prediction model, the apple cultivars were divided into 'old' representing heritage apple cultivars, introduced before *c.* 1835 when systematic selection efforts began, and 'new' including all apple cultivars introduced after this date. Only Malus x *domestica* cultivars were considered in this model, as the introduction date for the crab-apples was unknown and in

addition they have not been selected for consumption purposes. For the usage prediction model, the apple cultivars were divided into dessert, culinary, cider and ornamental. For apple cultivars of dual purpose, the primary use was included. For the harvest season prediction model, the apple cultivars were divided into Early (E) for cultivars harvested from late July to late August, Medium (M) for cultivars harvested between September to mid-October and Late (L) for cultivars harvested late October onwards. Individual models were constructed for a) phenolic compounds, b) sugars, c) phenolic compounds and sugars, both for dry weight (DW) and fresh weight (FW) concentrations. The FW and DW were obtained by weighing the samples before and after freeze-drying.

Steps involved in the models' calibration and validation are outlined in **Figure 1**. The total dataset was randomly divided into a training and a testing subset; consisting of 406 (75%) and 134 (25%) samples, respectively. Testing the models accuracy using a testing subset completely unknown to the developed models is far more indicative than the conventional leave-one-out-cross-validation method. Randomly dividing the dataset into training and testing subsets meant that in some cases replicates from the same cultivar were included in both sets, which could enhance performance accuracy. However, this was essential in order to avoid introducing bias by selectively excluding certain cultivars from the optimisation process as the model performance would then be dependable on the cultivars included within a particular training and testing distribution. Furthermore, in order to ensure the balance among the predicted classes (age, usage, and harvest season), a representative number of samples of each class were included in each subset. The training subset was then used to develop the classification models using the *k*NN, SVM, NB, and RF. For each classification approach, a grid search was performed in order to identify the most optimum parameter by examining the confusion matrix of the training dataset. The optimised models were then used for the models' calibration using the testing (unknown) subset created earlier. To assess the models stabilisation for each machine

learning technique applied, the previous steps were repeated as part of a 200 cycle process (100 cycles for the sugar models); where at each cycle, the training and testing subset samples were randomised and reshuffled.

In order to maximise the models' performance, different model input datasets were tested, including phenolic compounds and sugars alone or combined. The overall model performance for each classifier was assessed as a percentage value based on the total number of correct classification divided by the total number of samples within the testing subset.

RESULTS AND DISCUSSION

Phenolic profile of whole fruit and spatial distibution. The phenolic compounds identified for the 66 cultivars selected over three years for this study and their concentrations are summarised in **Table 2**. In addition, the qualitative and quantitative profiles of each separate tissue (peel, flesh, seeds) for 20 cultivars are presented in **Tables 3-5**. The final phenolic profile of each cultivar was influenced by the contribution of the peel, the flesh and the core (pericarp) which encloses the seeds. The relative contribution of each tissue to the biochemical profile of the whole apple depended on the size of each apple cultivar, which was very diverse between the different cultivars considered (**Supporting material S2**). Dessert cultivars usually had medium to large sized fruit, while most cider and ornamental apples, produced smaller-sized fruit. As a result, the phenolic composition of dessert apples was mainly influenced by the presence of the flesh (accounting for ~90% of the apples' weight); while in cider and ornamental cultivars the peel and the core had a greater influence per unit weight.

The profile of the whole apples was dominated by the presence of flavan-3-ols and oligomeric procyanidins, followed by hydroxycinammic acids, dihydrochalcones, quercetin glycosides and anthocyanins for the red cultivars. The collection of 22 apple cultivars was repeated over 2 or 3 years. A two-way Analysis of Variance (ANOVA) for 17 cultivars harvested in 2012 and

2013, indicated that cultivar x year interaction was significant for all the phenolic compounds ($p < 0.001$) considered (Supporting information **Table S4**). This result could be a reflection of the differences in agricultural practices and conditions as some cultivars were supplied from a commercial farm during 2012. Other factors influencing the yearly variation could include differences in maturity level and weather conditions. Despite some variation observed between years, the overall phenolic profile remained stable for most of the 22 cultivars assessed over more than one year, as supported by hierarchical cluster analysis. As shown in **Figure 2**, samples belonging to the same cultivar grouped together for ~80% of the cultivars showing that genotype is the major factor contributing to the phenolic profile. Other studies have also concluded that genotype was the most significant factor affecting primary and secondary metabolites in different apple cultivars followed by year.[12]

Among the different groups considered, ornamental apples exhibited the most interesting and unusual phenolic profile with the most characteristic trait being the high degree of pigmentation across peel, flesh and seeds. The degree of pigmentation in the flesh varied across different cultivars ranging from a pink tint to deep red/purple. 'Neville Copeman' and 'Brogdale Crab' were the only crab-apples studied with yellow-orange flesh. The highest cyanidin-3-O-galactoside (ideain) concentration recorded was 107.4 mg 100 g^{-1} FW for 'Royalty' a crab-apple bearing extremely small fruits. 'Royalty' is a very popular ever-red-leafed ornamental crab-apple producing deep red petals during the flowering period and red to purple fruits. 'Red Flesh', another crab-apple with an intense red colour in the peel and bright red flesh, also had high anthocyanin content with ideain concentrations of 88.9 mg 100 g^{-1} FW in the peel, 5.0 mg 100 g^{-1} FW in the flesh and 23.5 mg 100 g^{-1} FW in the seeds. These concentrations are more than 3-fold higher compared with other red apple cultivars and are more comparable to the levels found in highly pigmented soft fruits such as black currants (whole berry).[26]

The concentrations of the non-coloured phenolic compounds in ornamental apples were

genotype specific with different trends recorded across the different *Malus* species. Flavan-3-ols and procyanidins for instance were the dominant phenolic group in four of the eight cultivars examined, namely ‘Wisley Crab’, 'Niedzwetzkyana' Derivative, ‘Neville Copeman’, ‘Red Flesh’. (–)-Epicatechin and procyanidin B2 were the dominant flavonoids, with (+)-catechin and procyanidin B1 present in minor amounts, which is consistent with the profile of the majority of apple cultivars. ‘Wisley Crab’ showed a deviation from this pattern with 52.0 mg 100 g^{-1} FW (+)-catechin content, which is ~3-fold higher from other apple cultivars with high catechin levels. All other ornamental cultivars studied, exhibited the opposite trend with flavan-3-ols and procyanidins being present in very low to moderate amounts. The results from the peel and flesh also highlighted the unique characteristics of crab-apples and indicated that the elevated phenolic content observed for some ornamental cultivars was not merely a result of their small size. Crab-apples represented a rich source of dihydrochalcones and quercetin glycosides, while hydroxycinnamic acid levels varied. ‘Brogdale Crab’, ‘Neville Copeman’ and ‘Royalty’ in particular had the highest dihydrochalcone levels, with the latter having a mean phloridzin concentration of 100.7 mg 100 g^{-1} FW, which to the best of our knowledge is the greatest amount reported so far. A previous study,[27] also highlighted the fact that breeding material including popular commercial cultivars such as ‘Braeburn’ and ‘Golden Delicious’ appeared to have strongly reduced phenolic compounds compared to wild germplasm as assesed by total metabolite abundance. 'Niedzwetzkyana Derivative’ in particular, is a derivative of *Malus pumilla* Niedzwetzkyana, an endangered species native to Kazakhstan and Central Asia, which was brought in Europe in the late 19th century and is thought to be the main ancestor of most red-fleshed apples[28]. Indeed Nocker *et al.*,[29] who studied the genetic diversity among a wide selection of red-fleshed apples were able to trace the parentage of most accessions back to 'Niedzwetzkyana' with examples including ‘Red Flesh’ and ‘Maypole’. ‘Royalty’ was among the few red-fleshed apples which did not derive from 'Niedzwetzkyana'.

Crab-apples represent an understudied group of apple species which has only recently started to attract the interest of researchers especially the red-fleshed species which could be used as candidates for breeding programmes and a source for the development of nutraceuticals.[8,28–31] Moreover the vast number of wild species and hybrids available and their metabolic and phenotypic diversity makes the study of these apple species particularly interesting. Rudikovskaya[32] highlighted the unique phenolic profile of a Siberian crab-apple and its hybrids with domestic apples showcasing the potential for breeding new cultivars of apples with desired traits such as colour, disease resistance and health-promoting properties. Nevertheless, the phenolic profile of crab-apples is still largely unknown with only a few reports providing qualitative data.[8,30,33]

The other group of apples with significant phenolic content included cider cultivars with bittersweet cider cultivars in particular characterised by high amounts of flavan-3-ols, procyanidins and hydroxycinnamic acids. Although bitterness is usually undesirable in dessert apples, it is important for apples used to make beverages. Phenolic compounds have often been associated with bitter taste and astringency. It has been suggested that (–)-epicatechin, the main flavan-3-ol in apples, is more bitter than its stereoisomer (+)-catechin and that bitterness tends to diminish and astringency rises as the degree of polymerisation of procyanidins increases.[34] The highest phenolic contents were recorded for 'Pennard Bitter' and 'Stable Jersey', with the latter having approximately double the flavan-3-ol and hydroxycinnamic acid concentrations compared to the majority of other cider cultivars studied. The same trend was observed in the peel and the flesh of this cultivar, which contained extremely high levels of phenolic compounds, a result attributed to the genotype but also to the low maturity level of this cultivar as shown by the starch index (2.2) and the low sugar content (2.2 g 100 g^{-1} FW). Indeed, unripe apples have been shown to contain up to 10 times higher levels of phenolic compounds compared to apples harvested at optimum maturity.[35] Harvest at optimum maturity though is

not always the practice followed for some late season cider cultivars, which are often picked at windfall and left in storage for a few days to convert the starch into sugars before processing. Although the cider industry in the UK is among one of the biggest in the world, the phenolic content of British cider cultivars has been sparsely investigated, with the existing literature focusing mainly on the phenolic content of French and Spanish cider cultivars,[36,37] or on the beverage itself and the pomace which is the main by-product of the cider industry.[38]

The most extensive study of English cider apples examined the phenolic content of 19 cultivars (mainly bittersweets) and contributed in highlighting the variability in total phenolic content which ranged according to the different genotypes from 23 to 492 mg 100 g^{-1} FW in the flesh and 54.6 to 630.6 mg 100 g^{-1} FW in the peel.[39]

An important source of variation in the previous study was probably caused by horticultural maturity, since all cultivars were harvested at the same time and not at optimum maturity. This oversight could explain the elevated phenolic content of some cultivars compared to the results in the present study, in which all cider cultivars with the exception of 'Stable Jersey', were picked around the optimum harvest time. Despite this variation and the fact that the cultivars examined differed from the ones in the present study, (except 'Golden Delicious'), the results in both studies were in good agreement.

Culinary apples showed great variability in their phenolic content, with the sum of individual phenolic compounds ranging from 37.4 mg 100 g^{-1} FW for 'Beauty of Moray' to 176.7 mg 100 g^{-1} FW for 'Colonel Yate' over two successive years, with the latter having extremely high concentrations of chlorogenic acid and flavan-3-ols and procyanidins, similar to cider apples. The same profile was observed for 'Bramley's Seedling' the most commercially important culinary English apple cultivar, which had an average phenolic content of 90 mg 100 g^{-1} FW over two successive years.

Apart from culinary, ornamental and cider apples the majority of the cultivars considered in

this study included dessert apples as traditionally most apple cultivars were intended for fresh consumption. This group comprised a very diverse selection of modern mainstream cultivars including popular cultivars such as 'Gala', 'Braeburn' or the more recent 'Jazz', traditional heritage cultivars of UK origin such as 'Beauty of Bath', 'Cox's Orange Pippin', 'Devonshire Quarrenden' all important parent cultivars, and some very old or ancient cultivars including 'Old Pearmain (of Kelsey)', 'Decio' and 'Ribston Pippin'. The cultivars were broadly divided into 'old' and 'new' according to introduction date, meaning that some traditional English cultivars such as 'Egremont Russet' and 'Worcester Pearmain', were classified as 'new', although they have not arisen from intensive breeding programmes. In general, the results for the whole apple study showed that dessert cultivars had a lower phenolic content than cider apples, which is consistent with previous reports. 'Cox's Orange Pippin', an important progenitor for many modern cultivars, was among the apples with the lowest phenolic content (39.7 mg 100 g^{-1} FW), which is in agreement with previous reports for this cultivar.[40] The same pattern was observed for 'Queen Cox' a more highly pigmented clone of 'Cox's Orange Pippin'. The dominant phenolic compound in many dessert cultivars was chlorogenic acid, accounting for approximately 50% or more of the sum of phenolic compounds measured. High levels of chlorogenic acid are considered undesirable in a dessert apple as they are associated with extensive browning and bitterness. According to Ceymann and co-workers,[4] apple cultivars can be divided into hydroxycinnamic acid-dominated cultivars and flavan-3-ol-dominated cultivars depending on the ratio of flavan-3-ols and procyanidins / hydroxycinnamic acids. Based on this classification, most modern commercial apples had a balanced composition with ratios between 0.8 – 1.2. Less homogeneity was recorded within the very old and traditional apple group with only around 20% of them having a balanced ratio.

Apart from the concentration of chlorogenic acid, another important aspect considered in predicting the degree of browning for a particular cultivar, is the chlorogenic acid / *p*-

coumaroylquininc acid ratio. Chlorogenic acid is considered to be a preferential substrate of the catecholase activity of polyphenol oxidase (PPO), whereas *p*-coumaroylquininc acid is thought to be a competitive inhibitor of this enzyme activity.[41] Furthermore, PPO activity can be inhibited by the presence of procyanidins, and oxidation products of (–)-epicatechin,[41] therefore cultivars with high procyanidin / hydroxycinnamic acid ratio and low chlorogenic acid / *p*-coumaroylquinic acid ratio would be more suitable for the juice industry and fresh-cut fruit products. Cultivars fulfilling these criteria include 'D'Arcy Spice', 'Decio', 'Laxton Pioneer' and 'Queen Cox'.

Another group of phenolic compounds characteristic of apples is dihydrochalcones, with phloridzin being the most abundant followed by phloretin-2'-*O*-xylosylglucoside.[4] 'D'Arcy Spice' and 'Decio', two russet heritage cultivars deviated from this pattern, with phloretin-2'-*O*-xylosylglucoside present in approximately 1.5-fold higher concentrations than phloridzin and dihydrochalcones accounting for ~30% of the sum of phenolic compounds, compared to < 15% for most other dessert cultivars. The importance of dihydrochalcones in apples derives from the accumulating evidence for their potential health promoting properties, with phloridzin in particular showing promising potential for the regulation of blood sugar and prevention of type II diabetes.[10,11,14]

It is notable that the majority of dessert apples with high levels of dihydrochalcone were heritage or early modern cultivars, with examples including 'Wheeler's Russet', 'Devonshire Quarrenden' and 'Egremont Russet', a traditional English cultivar, which is still popular in the UK. Indeed, elevated phloridzin levels have been associated with the presence of russet in the skin, which is characteristic of several heritage cultivars and has gradually being bred out of most modern apple cultivars, based on a consumer preference for smooth highly coloured apples.[42] Also, many modern cultivars, such as 'Gala' and 'Ball's Pippin', have descended from 'Cox's Orange Pippin' a heritage cultivar particularly low in dihydrochalcones and other

phenolic groups. On the other hand, other important parent apple cultivars such as 'Devonshire Quarrenden' and 'Beauty of Bath', exhibited high levels of phenolic compounds which appear to have been inherited and passed on to some of their progeny, such as 'Worcester', 'Discovery', 'Redsleeves' and 'Scrumptious'. Earlier reports have indicated that heritage cultivars tend to have higher dihydrochalcone and hydroxycinnamic acid levels, compared to modern cultivars emerging from selection programmes.[15,42] The results of the present study although in accordance with this view, also highlight that this is not universal across all heritage cultivars. It can thus be concluded that high dihydrochalcone levels are specific to certain apple genotypes and are more prevalent among heritage cultivars.

Overall, the results from this study have highlighted the phytochemical variability among heritage cultivars some of which exhibit unique patterns of phenolics. In contrast, the majority of modern cultivars have a relatively similar phenolic profile. This may reflect that most modern cultivars have emerged from a restricted gene pool, as the number of progenitors used in commercial production over the last century has narrowed down to only a few genotypes. Volz and McGhie,[12] found that total peel and flesh phenolic compound concentrations were lower for genotypes originating in New Zealand which were derived primarily from 'Gala' (or it's mutant 'Royal Gala') and 'Braeburn', compared to genotypes originating outside New Zealand with dates of introduction spanning from 1600 to 1975. Furthermore, Khan and co-workers,[27] have found evidence that a strong reduction in phenolic compounds is possible within one breeding generation, known as negative transgressive segregation.

A more consistent trend in modern dessert cultivars was the greater levels of flavonols, in the form of quercetin glycosides, compared to heritage and traditional cultivars. This finding was supported by analysis of the peel and flesh tissues. The majority of quercetin glycosides was observed in the apple peel with all modern cultivars having a total concentration between 105.21 mg 100 g^{-1} FW for 'Scrumptious' to 167.28 mg 100 g^{-1} FW for 'Royal Gala'. Heritage

cultivars in contrast ranged between 9.01 mg 100 g^{-1} FW for 'Wheeler's Russet to 59.10 mg 100 g^{-1} FW for 'Old Pearmain (of Kelsey)'. Flesh flavonols were present in very low amounts and were under the quantification limit in most heritage cultivars. These results are in agreement with De Paepe and co-workers[15] who have recorded the same trend between classic/new and heritage cultivars. Other researchers in contrast have shown a poor correlation between flavonol content and genotype,[12,42] which is assumed to be caused by changes in environmental conditions across different years and geographical locations. These compounds are synthesised in response to ultra-violet (UV) light and variations in environmental temperature and have within a tree been shown to increase in apples growing in the outer tree canopy.[5] In the present study, a direct comparison could be attempted as all cultivars considered were grown on the same rootstocks and the majority of apple cultivars originated from the same NFC location. These observations were supported by HCA for the peel data from the 20-cultivar subset (**Figure 3A**) with quercetin glycosides contributing significantly to the separation. HCA also highlited the differences between different species with two major clusters formed, a large one containing the majority of Malus x *domestica* and hybrid cultivars and a smaller one containing mainly the wild crab-apples (**Figure 3)**. Good classification was also achieved based on genotype, with most biological replicates clustering together. In contrast, seeds exhibited a very different phenolic profile to the other tissues and were characterised by high biochemical variability within biological replicates. The main phenolic compounds in seeds was phloridzin followed by phloretin-2'-*O*-xylosylglucoside in much lower concentrations. Phloridzin concentrations varied as much as 10-fold between cultivars ranging from 290.4 mg 100 g^{-1} FW for 'Porter's Perfection' to 2113.5 mg 100 g^{-1} FW in 'Genet Moyle (of Taylor)'. Other phenolic groups present in apple seeds included chlorogenic acid, flavan-3-ols and dimeric procyanidins, and ideain present only in the seeds of ornamental apples with coloured seeds. Fromm and co-workers[43] have also reported similar levels of phenolic compounds in apple seeds from 12

different apple cultivars including 'Royal Gala'. They also detected and quantified several more hydroxycinnamic acids present in very low levels, as well as small amounts of 3-hydroxyphloretin.

Apple seeds are also characterised by the presence of the toxic cyanogenic glycoside amygdalin. Amygdalin levels varied between the twenty cultivars examined with the highest concentrations generally being observed in cider apples (**Table 5**). The concentrations of amygdalin detected in the seeds could generate from 0.95 to 9.6 mg cyanide equivalents per 100 g of apple seeds (FW), which is relatively high as the safe limits for humans are below 0.5 mg kg^{-1} body weight. Therefore, possible utilisation of apple seeds for the production of functional foods or nutraceuticals would require removal of amygdalin. In a previous study amygdalin content of 15 different apple cultivars was shown to vary from 95 mg 100 g^{-1} to 390 mg 100 g^{-1} of desiccated seeds.[44] These results were in good agreement with the concentrations measured in the present study with 'Egremont Russet' and 'Bramley' in particular having almost identical amygdalin concentrations after correcting for water content.

The present study seeked to emphasize the phenolic potential of the seeds while at the same time assessing the potential risks to human health. Earlier reports on apple seeds have concentrated on either the phenolic content[43] or the presence of amygdalin alone,[44] and again these only look at the seeds in isolation without considering the relationship with other tissues.

Soluble sugars and organic acids. The concentrations for soluble sugars for the 66 cultivars studied and organic acids for the 20-cultivar sub-sample studied are presented in Tables 6-7. The mean soluble sugar concentration for dessert apples was 9.4 g 100 g^{-1} FW compared to 7.8 g 100 g^{-1} FW for cider and 7.1 g 100 g^{-1} FW for ornamental cultivars. Fructose was dominant across all cultivars accounting for ~50% of total sugars, followed by sucrose with a ~30% contribution which is in line with previous reports.[45,46] Glucose varied considerably across cultivars, accounting between 6.6% for 'D'Arcy Spice' to 28.5% of total sugars for 'Cummy

Norman'. Apples were also characterised by the presence of small to moderate amounts of sorbitol, a sugar alcohol which along with sucrose is the end product of photosynthesis in the leaves and is then transported to other parts of the plant, such as fruits and seeds.[47] Sorbitol accounted for between 3.3 to 8.0% of the total sugar content of apples. Culinary apples had the lowest mean sugar content (7.72 mg 100 g^{-1} FW), although this was genotype dependent with some cultivars like 'Colonel Yate' having similar levels to desert apples. The sugar content measured was also a reflection of the maturity stage of each cultivar when harvested. Although care was taken to select apples of optimum commercial maturity, there was some variation, which was more obvious for ornamental and cider apples, as can be seen in **Table S2**. This can be attributed to several reasons such as subjective evaluation of picking times, especially for ornamental apples, weather conditions and climatic differences between years.

Three different organic acids were detected in 20 apple cultivars, with malic acid being dominant, followed by moderate amounts of quinic acid and ascorbic acid, which was detected in less than 50% of the samples in very small amounts up to 5.5 mg 100 g^{-1} FW (**Table 7**). Malic acid levels varied considerably between different cultivars and within groups, ranging between 287.0 to 2858.7 mg 100 g^{-1} FW for 'Cummy Norman' and 'Maypole', respectively. Cultivars used for culinary purposes such as 'Bramley' and 'Gennet Moyle (of Taylor)' a dual purpose apple cultivar, showed a high malic acid concentration, which may contribute to the characteristic tart taste to these apples. Cider apples have been selected to have a broad spectrum of organoleptic properties with different ratios of sugars, organic acids and tannins giving cider their characteristic taste ('hard', 'sharp', 'sweet', bittersweet, 'bitter-sharp'). For example, 'Cummy Norman' a bittersweet cider cultivar was characterised by low organic acids, high sugar and high phenolic content, especially flavan-3-ols, procyanidins and chlorogenic acid which are more related with the development of bitter taste. Phloridzin is also a compound with a bitter taste[13] which was found in elevated concentrations in cider cultivars. High sugar is also

essential to balance off the bitter taste while alcohol has been suggested to enhance it.[48]

Apples can contain a variety of other organic acids including citric, fumaric, shikimic and maleic acid in the range of 10-100 μg g^{-1} (fresh weight) which rapidly decline during maturation as has been demonstrated by previous studies.[48] The apples in the present study were collected at optimum maturity which can account for the fact that no other organic acids were detected. The levels of organic acids and sugars are in agreement with previous reports, considering primarily dessert apple cultivars.[46,49,50]

Classification models to predict age, usage and harvest season. The biochemical profile of the whole apple extracts was further used to develop a series of prediction models using *k*NN, SVM, NB, and RF. The prediction accuracies of the different models developed based on the phenolic and sugar data sets are displayed in **Figure 4**. The dataset including both phenolic compounds and sugars had a slightly worse performance than the phenolic dataset and has not been included. Among the different machine learning methods applied, RF consistently achieved the best validation performance for all the different combinations, with a prediction accuracy ≥ 90% for all three models based on their phenolic profile with the highest success rate observed for the usage model (95.5%). Moreover, the confusion matrix for the prediction of usage based on the phenolic profile, showed that 100% of cider and dessert apples were correctly classified (true positives), while three ornamental cultivars were classified as dessert including the hybrids 'Maypole' and Krasnyi Shtandart' which are dual purpose. In addition, approximately 31% of the culinary samples were also classified as either dessert or cider reflecting the dual purpose of these cultivars.

SVM and *k*NN scored > 80% for all three models based on the phenolic profile, while NB exhibited the lowest prediction accuracy of approximately 70% for usage and age, while for harvest season it failed to accurately classify more than 50% of the samples.

The RF models constructed with the sugar dataset achieved prediction accuracy between approximately 70-80%, with the best results observed for the prediction of usage, with glucose and fructose mainly driving the separation. The high success rate for the prediction of usage observed with both the phenolic and sugar datasets could reflect their role in the organoleptic properties which determine the suitability of apple genotypes for different uses.

Next, a decision tree was generated based on the best performance models in order to identify the key compounds contributing to the observed classification (**Figure 5**). Procyanidin B2, ideain, and *p*-coumaroylquinic acid were the phenolic compounds contributing the most to the dataset variance for the prediction of usage, while procyanidin B1, (+)-catechin and *p*-coumaroylquinic acid were the phenolic compounds contributing to the input dataset variance for the prediction of age, despite being present in the apples in relatively low concentrations. Among the sugars, glucose was the variable contributing most to the separation, followed by fructose (data not shown), which is consistent with the observation in the previous section, that glucose exhibited higher variability between cultivars. These findings can help understand the differences between apples intended for different uses and guide future breeding programmes. Furthermore, the harvest season model was influenced by the presence of ideain, chlorogenic acid, procyanidin B2 and phloridzin. High ideain levels in particular, were associated with ornamental apples which mostly ripen during the early season. On the other hand, late season cultivars were associated with high chlorogenic, procyanidin B2 and phloridzin levels which is characteristic of most cider cultivars.

To our knowledge, this is the first time machine learning techniques have been employed to predict certain traits of apple cultivars based on their biochemical profile. Moreover, this is the first time the sugar profile has been successfully used to develop prediction models for the classification of apple cultivars according to usage, age and harvest season, showcasing the advantages of using machine learning methods over traditional statistically based classification

methods. The high success rate in differentiating between 'new' and 'old' cultivars further supports the hypothesis that cultivars introduced before the intensification of breeding programmes have a distinct biochemical profile and there may be scope revisiting some of them in an effort to develop food products with enhanced health promoting properties.

These results highlight the potential of machine learning for mapping the metabolic profile of a large collection of apple cultivars intended for very diverse uses and introduction dates spanning over several centuries. Knowledge of the key metabolites contributing to flavour and/or health-promoting properties could guide cultivar selection in future breeding programmes on the basis of phytochemical content.

AUTHOR INFORMATION

Corresponding Author

*(L.A.T.) Phone: +44(0)1234750111 E-mail:l.a.terry@cranfield.ac.uk

Funding

Authors M. Anastasiadi, L. A. Terry, S. P. Redfern, M. Berry and M. S. J. Simmonds received funding from Innovate UK Grant TSB 101125, Unilever R&D, and the Biotechnology and Biological Sciences Research Council (BBSRC).

ACKNOWLEDGMENTS

The authors are grateful to Dr M. Ordidge from the University of Reading for providing access to the National Fruit Collection and Farm Advisory Services Team (FAST) for helping with the sample collection. Also to Mr. A. Harding from WorldWide Fruit Ltd. for supplying some of the commercial apple cultivars and to Dr S. Foukaraki for helping with the analysis of non-

volatile organic acids. The underlying data can be accessed at https://doi.org/10.17862/cranfield.rd.5002442.

ASSOCIATED CONTENT

Supporting Information Available: Firmness, skin colour, starch content and fruit dimensions of the apple cultivars used in the study. Log10 transformed data for phenolic composition of whole apples, Fisher's F-values and significance associated with ANOVAs (cultivar, year) performed on phenolic compounds for the cultivars harvested in 2012 - 2013.

This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

(1) Forsline, P. L.; Aldwinckle, H. S.; Dickson, E. E.; Luby, J. J. Collection, Maintainance, Characterization, and Utilization of Wild Apples of Central Asia. In *Horticultural Reviews*, 1; Janick, J.; Ed.; Wild Apple and Fruit Trees of Central Asia. John Wiley & Sons, New York, 2003; 29, 1-304.

(2) Planchon, V.; Lateur, M.; Dupont, P.; Lognay, G. Ascorbic acid level of Belgian apple genetic resources. *Sci. Hortic.* **2004,** *100*, 51–61.

(3) Vrhovsek, U.; Rigo, A.; Tonon, D.; Mattivi, F. Quantitation of Polyphenols in Different Apple Varieties. *J. Agric. Food Chem.* **2004,** *52*, 6532–6538.

(4) Ceymann, M.; Arrigoni, E.; Schärer; H., Bozzi Nising, A.; Hurrell, R. F. Identification of apples rich in health-promoting flavan-3-ols and phenolic acids by measuring the polyphenol profile. *J. Food Compos. Anal.* **2012,** *26*, 128–135.

(5) Watkins, C. B.; Liu, R. H. Pome Fruit. In *Health-Promoting Properties of Fruits and Vegetables,* 1; Terry, L. A., Ed.; CABI, Oxford, UK, 2011; 196-217.

(6) Velasco, R.; Zharkikh, A.; Affourtit, J.; Dhingra, A.; Cestaro, A.; Kalyanaraman, A.; Fontana, P.; Bhatnagar, S. K.; Troggio, M.; Pruss, D.; et al. The genome of the domesticated apple (*Malus* × *domestica* Borkh.). *Nat. Genet.* **2010,** *42*, 833–839.

(7) Cornille, A.; Giraud, T.; Smulders, M. J. M.; Roldán-Ruiz, I.; Gladieux, P. The domestication and evolutionary ecology of apples. *Trends Genet.* **2014,** *30*, 57–65.

(8) Shen, H.; Zhang, J.; Yao, Y.; Tian, J.; Song, T.; Geng, J.; Gao, J. Isolation and expression of McF3H gene in the leaves of crabapple. *Acta Physiol. Plant.* **2012,** *34*, 1353–1361.

(9) Morgan, J.; Richards, A. *The New Book of Apples: The Definitive Guide to Apples, Including Over 2000 Varieties.* 2; Ebury Press, London, UK; 2002; 137-163.

(10) Ehrenkranz, J. R. L.; , Lewis, N. G.; Ronald Kahn, C.; Roth, J. Phlorizin: a review. *Diabetes Metab. Res. Rev.* **2005,** *21*, 31–38.

(11) Masumoto, S.; Akimoto, Y.; Oike, H.; Kobori, M. Dietary Phloridzin Reduces Blood Glucose Levels and Reverses Sglt1 Expression in the Small Intestine in Streptozotocin-Induced Diabetic Mice. *J. Agric. Food Chem.* **2009,** *57*, 4651–4656.

(12) Volz, R. K.; McGhie, T. K. Genetic Variability in Apple Fruit Polyphenol Composition in *Malus* × *domestica* and *Malus sieversii* Germplasm Grown in New Zealand. *J. Agric. Food Chem.* **2011,** *59*, 11509–11521.

(13) Gosch, C.; Halbwirth, H.; Stich, K. Phloridzin: Biosynthesis, distribution and physiological relevance in plants. *Phytochemistry* **2010,** *71*, 838–843.

(14) Makarova, E.; Górnaś, P.; Kornade, I.; Tirzite, D.; Cirule, E.; Gulbe, A.; Pugajeva, I.; Seglina, D.; Dambrova, M. Acute anti-hyperglycaemic effects of an unripe apple preparation containing phlorizin in healthy volunteers: a preliminary study. *J. Sci. Food Agric.* **2015,** *95*, 560–568.

(15) Paepe, D. D.; Valkenborg, D.; Noten, B.; Servaes, K.; Diels, L.; Loose, M. B.; Droogenbroeck, B., V.; Voorspoels, S. Variability of the phenolic profiles in the fruits from old, recent and new apple cultivars cultivated in Belgium. *Metabolomics* **2014,** *11*, 739–752.

(16) Mohareb, F.; Papadopoulou, O.; Panagou, E.; Nychas, G.-J.; Bessant, C. Ensemble-based support vector machine classifiers as an efficient tool for quality assessment of beef fillets from electronic nose data. *Anal. Methods* **2016**, *8*, 3711–3721.

(17) Panagou, E. Z.; Mohareb, F. R.; Argyri, A. A.; Bessant, C. M.; Nychas, G.-J. E. A comparison of artificial neural networks and partial least squares modelling for the rapid detection of the microbial spoilage of beef fillets based on Fourier transform infrared spectral fingerprints. *Food Microbiol*. **2011**, *28*, 782–790.

(18) Jeong, J. H.; Resop, J. P.; Mueller, N. D.; Fleisher, D. H.; Yun, K.; Butler, E. E.; Timlin, D. J.; Shim, K.-M.; Gerber, J. S.; Reddy, V. R.; Kim, S.-H. Random Forests for Global and Regional Crop Yield Predictions. *PLOS ONE* **2016**, *11*, e0156571.

(19) Porfirio, L. L.; Newth, D.; Harman, I. N.; Finnigan, J. J.; Cai, Y. Patterns of crop cover under future climates. *Ambio* **2017**, *46*, 265–276.

(20) Anastasiadi, M.; Mwangi, P. M.; Ordaz-Ortiz, J. J.; Redfern, S. P.; Berry, M.; Simmonds, M. S. J.; Terry, L. A. Tissue biochemical diversity of 20 gooseberry cultivars and the effect of ethylene supplementation on postharvest life. *Postharvest Biol. Technol.* **2016,** *117*, 141–151.

(21) Silverman, B. W.; Jones, M. C. E. Fix and J.L. Hodges (1951): An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation: Commentary on Fix and Hodges (1951). *Int. Stat. Rev.* **1989**, *57*, 233–238.

(22) Beygelzimer, A.; Kakadet, S.; Langford, J.; Arya S.; Mount, D. FNN: Fast Nearest Neighbor Search Algorithms and Applications. **2015**, URL (https://cran.r-project.org/package=FNN) (Accessed: 16th April 2017).

(23) Breiman, L. Random Forests. *Mach. Learn*. **2001**, 45, 5–32.

(24) Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2* (3), 18-22.

(25) Dimitriadou, E.; Hornik, K.; Leisch, F.; Meyer D.; Weingessel, A. E1071: Misc Functions of the Department of Statistics (E1071), TU Wien. **2011.** URL (http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.191.317&rep=rep1&type=pdf) (Accessed Apr 17, 2017).

(26) Bordonaba, J. G.; Terry, L. A. Biochemical Profiling and Chemometric Analysis of Seventeen UK-Grown Black Currant Cultivars. *J. Agric. Food Chem.* **2008,** *56*, 7422–7430.

(27) Khan, S. A.; Tikunov, Y.; Chibon, P.-Y.; Maliepaard, C.; Beekwilder, J.; Jacobsen, E.; Schouten, H. J. Metabolic diversity in apple germplasm. *Plant Breed*. **2014,** *133*, 281–290.

(28) A. D. Dzhangaliev, T. N. S.; Salova, T. N.; Turekhanova, P. M. The Wild Fruit and Nut Plants of Kazakhstan. In *Horticultural Reviews*, 1; Janick, J.; Ed.; Wild Apple and Fruit Trees of Central Asia. John Wiley & Sons, New York, 2003; 29, 305-371.

(29) Nocker, S.; Berry, G.; Najdowski, J.; Michelutti, R.; Luffman, M.; Forsline, P.; Alsmairat, N.; Beaudry, R.; Nair, M. G.; Ordidge, M. Genetic diversity of red-fleshed apples (*Malus*). *Euphytica* **2011,** *185*, 281–293.

(30) Seeram, N. P.; Cichewicz, R. H.; Chandra, A.; Nair, M. G. Cyclooxygenase Inhibitory and Antioxidant Compounds from Crabapple Fruits. *J. Agric. Food Chem.* **2003,** *51*, 1948–1951.

(31) Wang, X.; Li, C.; Liang, D.; Zou, Y.; Li, P.; Ma, F.; Phenolic compounds and antioxidant activity in red-fleshed apples. *J. Funct. Foods* **2015,** *18*, Part B, 1086–1094.

(32) Rudikovskaya, E. G. Peculiarities of polyphenolic profile of fruits of Siberian crab apple and its hybrids with *Malus x Domestica* Borkh. *Acta Physiol. Plant.* **2015,** *37*, 37:238.

(33) Sommella, E.; Pepe, G.; Pagano, F.; Ostacolo, C.; Tenore, G. C.; Russo, M. T.; Novellino, E.; Manfra, M.; Campiglia, P. Detailed polyphenolic profiling of Annurca apple (M. pumila Miller cv Annurca) by a combination of RP-UHPLC and HILIC, both hyphenated to IT-TOF mass spectrometry. *Food Res. Int.* **2015,** *76*, Part 3, 466–477.

(34) Peleg, H.; Gacon, K.; Schlich, P.; Noble, A. C. Bitterness and astringency of flavan-3-ol monomers, dimers and trimers. *J. Sci. Food Agric.* **1999,** *79*, 1123–1128.

(35) Zheng, H.-Z.; Kim, Y.-I.; Chung, S.-K. A profile of physicochemical and antioxidant changes during fruit growth for the utilisation of unripe apples. *Food Chem.* **2012**, 131, 106–110.

(36) Le Deun, E.; Van der Werf, R.; Le Bail; G., Le Quéré; J.-M.; Guyot, S. HPLC-DAD-MS Profiling of Polyphenols Responsible for the Yellow-Orange Color in Apple Juices of Different French Cider Apple Varieties. *J. Agric. Food Chem.* **2015,** *63*, 7675–7684.

(37) Del Campo, G.; Berregi; I., Iturriza, N.; Santos, J. I. Ripening and Changes in Chemical Composition of Seven Cider Apple Varieties. *Food Sci. Technol. Int.* **2006,** *12*, 477–487.

(38) Diñeiro García, Y.; Valles, B. S.; Picinelli Lobo, A. Phenolic and antioxidant composition of by-products from the cider industry: Apple pomace. *Food Chem.* **2009,** *117*, 731–738.

(39) Marks, S. C.; Mullen, W.;Crozier, A. Flavonoid and chlorogenic acid profiles of English cider apples. *J. Sci. Food Agric.* **2007,** *87*, 719–728.

(40) Wojdyło, A.; Oszmiański, J.; Laskowski, P. Polyphenolic Compounds and Antioxidant Activity of New and Old Apple Varieties. *J. Agric. Food Chem.* **2008,** *56*, 6520–6530.

(41) Le Bourvellec, C.; Le Quéré, J.-M.; Sanoner, P.; Drilleau, J.-F.; Guyot, S. Inhibition of Apple Polyphenol Oxidase Activity by Procyanidins and Polyphenol Oxidation Products. *J. Agric. Food Chem.* **2004,** *52*, 122–130.

(42) Andre, C. M.; Greenwood, J. M.; Walker, E. G.; Rassam, M.; Sullivan, M.; Evers, D.; Perry, N. B.; Laing, W. A. Anti-Inflammatory Procyanidins and Triterpenes in 109 Apple Varieties. *J. Agric. Food Chem.* **2012,** *60*, 10546–10554.

(43) Fromm, M.; Bayha, S.; Carle, R.; Kammerer, D. R. Characterization and Quantitation of Low and High Molecular Weight Phenolic Compounds in Apple Seeds. *J. Agric. Food Chem.* **2012,** *60*, 1232–1242.

(44) Bolarinwa, I. F.; Orfila, C.; Morgan, M. R. A. Determination of amygdalin in apple seeds, fresh apples and processed apple juices. *Food Chem.* **2015,** 170, 437–442.

(45) Suni, M., Nyman; M., Eriksson; N.-A., Björk, L.; Björck, I. Carbohydrate composition and content of organic acids in fresh and stored apples. *J. Sci. Food Agric.* **2000,** *80*, 1538–1544.

(46) Wu, J.; Gao, H.; Zhao, L.; Liao, X.; Chen, F.; Wang, Z.; Hu, X. Chemical compositional characterization of some apple cultivars. *Food Chem.* **2007,** *103*, 88–93.

(47) Beck, E.; Hopf, H. Branched-chain Sugars and Sugar Alcohols. In *Methods in plant biochemistry*. *Carbohydrates*, 1; Dey, P. M.; Harborne, J. B.; Ed.; Academic Press, London, UK, 1990; 2, 235-290.

(48) Fischer, U.; Noble, A. C. The Effect of Ethanol, Catechin Concentration, and pH on Sourness and Bitterness of Wine. *Am. J. Enol. Vitic.* **1994,** *45*, 6–10.

(49) Zhang, Y., Li, P.; Cheng, L. Developmental changes of carbohydrates, organic acids, amino acids, and phenolic compounds in 'Honeycrisp' apple flesh. *Food Chem.* **2010,** *123*, 1013–1018.

(50) Drogoudi, P. D.; Michailidis, Z.; Pantelidis, G. Peel and flesh antioxidant content and harvest quality characteristics of seven apple cultivars. *Sci. Hortic.* **2008,** *115*, 149–153.

FIGURE CAPTIONS

Figure 1. Flowchart illustrating the construction and optimisation process for the development of prediction models for age, usage and harvest season using *k*NN, SVM, NB, and RF.

Figure 2 HCA constellation plot showing clustering of 22 apple cultivars harvested in different years based on their phenolic profile. For explanation of cultivar name and phenolic compound abbreviations refer to **Table 1** and **Table 2** respectively.

Figure 3. HCA and Heat maps for the peel (A) and flesh (B) phenolic compound datasets**.** The colour code provided prepresents the concentrations of each invididual compound in ascending order from green to red. For explanation of cultivar name and phenolic compound abbreviations refer to **Table 1** and **Table 2** respectively.

Figure 4. Overall prediction accuracy over 200 iterations (phenolic compounds) and 100 iterations (sugars) of different methods used to classify apple cultivars according to Usage, Age and Harvest Season. (A)-(B) corresponds to Usage prediction accuracy for phenolics and sugars respectively, (C)-(D) corresponds to Age prediction accuracy phenolics and sugars respectively, (E)-(F) corresponds to Harvest Season prediction accuracy for phenolics and sugars respectively.

Figure 5. Decision trees describing the contribution of each individual phenolic compound to the classification of apple cultivars according to Usage (A), Age (B) and Harvest Season (C). After each decision step, the original dataset is split in two smaller subsets based on the levels of each phenolic compound and each subset is assigned to a new class. The numbers inside each box (scale 0-1), represent what percentage of samples belonging to each class remains within a subset after each step. The cumulative success rate can be extracted at the final step by

taking into account the percentage of correctly classified samples for each subset. For explanation of phenolic compound abbreviations refer to **Table 2.**

Tables

Table 1. Information on apple cultivars collected[a].

Species	Cultivar	Acronym	Usage	Harvest season	Introduction date	Age	Country of origin	Parentage	Years assessed
M. x domestica	'Captain Broad'	CBR	cider	L	undated	OLD	UK	unknown	2012-2013
M. x domestica	'Porter's Perfection'	PPRF	cider	L	c. 1800s	OLD	UK	unknown	2012-2014
M. x domestica	'Cummy Norman	CUMN	cider	M	undated	OLD	UK	unknown	2012, 2014
M. x domestica	'Stable Jersey'	STJR	cider (bittersweet)	L	undated	OLD	UK	unknown	2012
M. x domestica	'Dymock Red'	DYMR	cider (bittersweet)	E	undated	OLD	UK	unknown	2012, 2014
M. x domestica	'Pennard Bitter'	PENB	cider (bittersweet)	L	undated	OLD	UK	unknown	2012
M. x domestica	'Sops-in-Wine' (1992-133)	SIW92	cider/culinary	M	undated	OLD	UK	unknown	2012
M. x domestica	'Sops-in-Wine' (1979-036)	SIW79	cider/culinary	M	c. 1800	OLD	UK	unknown	2012
M. x domestica	'Gennet Moyle (of Taylor)'	GMT	cider/culinary	M	c. 1600s	OLD	UK	unknown	2012-2013
M. x domestica	'Golden Spire'	GSPR	culinary/cider	E	1850	NEW	UK	unknown	2012
M. x domestica	'Keswick Codlin'	KCODL	culinary	E	1793	OLD	UK	unknown	2012
M. x domestica	'Bramley's Seedling'	BRAM	culinary	M	c. 1809	OLD	UK	unknown	2012-2013
M. x domestica	'Charlotte'	CHARL	culinary	L	1975	NEW	UK	unknown	2012
M. x domestica	'Beauty of Moray'	BMOR	culinary	E	1883	NEW	UK	unknown	2012
M. x domestica	'Domino'	DOM	culinary	E	1883	NEW	UK	unknown	2012-2013

M. x domestica	‘Colonel Yate'	COLY	culinary	L	1905	NEW	UK	Lane's Prince Albert x Peasgood's Nonsuch	2012-2013
M. x domestica	‘Bedfordshire Foundling'	BEDSF	culinary	L	1800	OLD	UK	unknown	2012
M. x domestica	‘Ashmeads Kernel'	ASMK	dessert	L	1700	OLD	UK	unknown	2012
M. x domestica	‘Ball's Pippin'	BALPN	dessert	L	1923	NEW	UK	Cox's Orange Pippin x Stummer Pippin	2012
M. x domestica	‘Beauty of Bath'	BBATH	dessert	E	1864	NEW	UK	unknown	2013
M. x domestica	‘Beauty of Bedford'	BBED	dessert	M	1913	NEW	UK	Lady Sudely x Beauty of Bath	2012
M. x domestica	‘Black McIntosh'	BMCI	dessert	M	1928	NEW	Canada	Sport of McIntosh (Fameuse x Unknown)	2012
M. x domestica	‘Braeburn'	BRAEB	dessert	L	1950	NEW	New Zealand	Lady Hamilton x unknown	2012
M. x domestica	‘Cambusnethan Pippin'	CAMBP	dessert/culinary	M	c. 1750	OLD	UK (Scotland)	unknown	2012
M. x domestica	‘Christmas Pearmain'	CHRSP	dessert	L	1893	NEW	UK	unknown	2012
M. x domestica	‘Cox's Orange Pippin'	COP	dessert	L	c. 1825	OLD	UK	Ribston Pippin x unknown	2012
M. x domestica	‘D'Arcy Spice'	DAS	dessert	L	1785	OLD	UK	unknown	2012-2014
M. x domestica	‘Decio'	DEC	dessert	L	c. 450	OLD	Italy	unknown	2012
M. x domestica	Devonshire Quarrenden'	DEVQR	dessert	E	c. 1678	OLD	UK	unknown	2013-2014
M. x domestica	‘Discovery'	DISC	dessert	E	1949	NEW	UK	Worcester x Beauty of Bath	2012-2013

M. x domestica	'Duchess's Favourite'	DUCHF	dessert	E	c. 1800	OLD	UK	unknown	2012
M. x domestica	'Early Windsor'	EWIND	dessert	E	1930	NEW	Germany	Cox's Orange Pippin	2012
M. x domestica	'Egremont Russet'	EGR	dessert	M	c. 1872	NEW	UK	unknown	2012-2014
M. x domestica	'Elton Beauty'	ELTB	dessert	M	1952	NEW	UK	James Grieve x Worcester	2012
M. x domestica	'Epicure'	EPIC	dessert	E	1929	NEW	UK	Wealthy x Cox's Orange Pippin	2012
M. x domestica	'Falstaff'	FALS	dessert	L	1966	NEW	UK	James Grieve x Golden Delicious	2012
M. x domestica	'Gala'	GALA	dessert	L	1934	NEW	New Zealand	Kidds Orange Red x Golden Delicious	2013
M. x domestica	'Golden Delicious'	GDEL	dessert	L	1890	NEW	USA	unknown	2012-2013
M. x domestica	'Golden Knob'	GKNOB	dessert	L	late 1700s	OLD	UK	unknown	2012
M. x domestica	'Golden Pippin'	GPIP	dessert	L	c. 1629	OLD	UK	unknown	2012
M. x domestica	'Histon Favourite'	HISTF	dessert	M	1883	NEW	UK	unknown	2012-2013
M. x domestica	'James Grieve'	JGR	dessert	E	1893	NEW	UK (Scotland)	Pott's Seedling or Cox's Orange Pippin x Unknown	2013
M. x domestica	'Jazz'	JAZZ	dessert	M	2000	NEW	New Zealand	Braeburn x Royal Gala	2012
M. x domestica	'Lady Lambourne'	LADYL	dessert	M	1945	NEW	UK	Sport of Lord Lambourne	2012
M. x domestica	'Laxton Pioneer'	LAXP	dessert	M	1934	NEW	UK	Cox's Orange Pippin x Worcester	2012

M. x domestica	‘Lodgemore Nonpareil'	LODGN	dessert	M	1808	OLD	UK	unknown	2012
M. x domestica	‘Lord Lambourne'	LORDL	dessert	M	1907	NEW	UK	James Grieve x Worcester	2012
M. x domestica	‘Old Pearmain (of Kelsey)	OPMK	dessert	L	c. 1200s	OLD	UK or France	unknown	2012
M. x domestica	‘Pomme Noire'	PNOIR	dessert	L	undated	OLD	France	unknown	2012
M. x domestica	‘Queen Cox'	QCOX	dessert	L	1953	NEW	UK	Clone of Cox's Orange Pippin	2012
M. x domestica	‘Redsleeves'	REDSL	dessert	E	1986	NEW	UK	Exeter Cross x scab resistant seedling.	2012-2013
M. x domestica	‘Ribston Pippin'	RIBP	dessert	L	1707	OLD	UK	unknown	2012-2013
M. x domestica	‘Royal Gala	RGALA	dessert	L	1934	NEW	New Zealand	Sport of Gala	2012
M. x domestica	‘Scrumptious'	SCRUMP	dessert	E	1985	NEW	UK	Starkspur Golden Delicious x Discovery	2012-2013
M. x domestica	‘Spartan'	SPRTN	dessert	L	1926	NEW	Canada	McIntosh x Yellow Newton	2012
M. x domestica	‘Thorle Pippin'	THORLP	dessert	E	c. 1800	OLD	UK (Scotland)	unknown	2012
M. x domestica	‘Wheeler's Russet'	WHLR	dessert	L	1717	OLD	UK	unknown	2012, 2014
M. x domestica	‘Worcester Pearmain'	WRP	dessert	E	1874	NEW	UK	Devonshire Quarrenden x unknown	2012-2013
hybrid	‘Krasnyi Shtandart'	KRAST	ornamental/dessert	M	1915	NEW	Russia	Pepin Shafrannyi x Rubynovoye	2012

hybrid	‘Maypole'	MAYP	ornamental/culinary	M	1976	NEW	UK	Wijcik × *Malus* Baskatong	2012
Malus	‘Royalty	ROYAL	ornamental	E	no data	unknown	unknown	unknown	2012
Malus	‘Wisley Crab'	WSLCB	ornamental	E	no data	unknown	unknown	unknown	2012
Malus	‘Neville Copeman'	NVCP	ornamental	E	no data	unknown	unknown	unknown	2012, 2014
Malus	‘Red Flesh'	RDFL	ornamental	E	no data	unknown	unknown	unknown	2012-2013
Malus	‘Niedzwetzkyana' Derivative	NDER	ornamental	E	undated	unknown	unknown	unknown	2012-2013
Malus	‘Brogdale Crab'	BRGCB	ornamental	E	c. 1831	OLD	France	unknown	2013

[a] data have been compiled from the NFC database

Table 2 Phenolic composition of whole apples expressed in mg 100 g^{-1} FW[a]

Cultivar	CAT[b]	EPIC	PROCB1	PROCB2	CA	PQCA	PHLOR	PHLXG	QGAL	QGLU	QUER	QXYL	AVIC	IDEAIN	ΣPC
CBR 2012	3.00	18.16	2.16	36.16	62.99	4.31	8.73	9.14	3.64	1.55	1.32	1.01	2.74	0.03	154.9
CBR 2013	2.55	16.67	2.04	33.67	59.97	3.81	6.24	8.27	3.59	1.52	1.54	1.31	1.90	0.03	143.1
CUMN 2012	8.34	27.50	6.23	28.67	61.86	2.84	15.30	6.47	6.72	1.48	3.36	1.01	3.61	0.10	173.5
CUMN 2014	6.23	23.81	4.62	21.45	54.49	2.62	14.15	5.98	3.73	1.29	2.59	0.78	2.69	0.14	144.6
PPRF 2012	5.63	20.16	4.83	40.58	52.83	10.81	11.98	2.66	5.83	0.86	1.60	0.57	1.74	1.69	161.8
PPRF 2013	7.56	17.17	6.02	39.12	66.22	12.69	19.38	4.32	5.29	1.02	1.69	1.02	1.93	1.06	184.5
PPRF 2014	7.29	18.14	5.32	30.90	58.28	11.05	15.85	3.18	7.52	2.55	1.96	0.80	2.20	0.47	165.5
DYMR 2012	3.36	15.37	2.49	33.85	53.44	6.87	26.27	2.40	2.71	2.36	0.75	0.45	1.71	0.88	152.9
DYMR 2014	2.98	13.89	2.40	23.59	55.17	5.74	22.40	1.97	2.02	2.28	0.70	0.43	1.48	1.00	136.0
PENB	8.61	44.95	7.87	45.36	80.09	8.11	12.14	9.16	5.49	1.99	0.80	0.64	2.08	0.03	227.3
STJR	6.99	71.96	5.67	80.91	118.30	0.05	28.79	5.19	4.26	2.22	2.42	0.93	2.34	0.38	330.4
GMT 2012	2.94	11.32	2.12	21.82	70.05	0.48	7.32	5.69	2.92	1.11	1.06	0.39	1.28	0.25	128.8
GMT 2013	2.08	6.94	1.54	16.67	58.00	0.33	9.04	8.07	3.62	1.41	1.17	0.77	1.48	0.41	111.5
SIW79	1.83	8.24	1.33	14.12	24.54	4.38	6.24	2.63	4.02	0.54	1.25	1.10	3.20	2.17	75.6
SIW92	2.77	11.40	1.68	16.93	24.99	5.74	6.44	1.74	4.17	0.65	1.31	1.10	2.61	0.99	82.5
GSPR	1.05	7.55	0.72	11.69	25.59	1.19	5.06	0.69	2.26	0.35	1.63	0.48	1.29	0.03	59.6
BMOR	0.09	1.28	0.11	0.10	25.10	1.62	4.41	2.38	0.43	0.14	0.73	0.20	0.83	0.03	37.4
BRAM 2012	8.24	17.09	4.24	19.30	41.91	2.63	4.31	3.64	0.91	0.65	0.70	0.26	0.96	0.03	104.9
BRAM 2013	4.48	9.40	3.31	14.57	31.24	1.80	3.75	4.56	1.09	0.41	0.77	0.34	0.76	0.03	76.5
CHARL	1.21	6.45	0.93	7.78	14.01	0.94	4.25	1.92	0.97	0.45	0.72	0.31	0.82	0.18	40.9
COLY 2012	5.84	23.85	4.52	41.42	89.89	1.93	4.54	2.92	2.19	1.10	0.86	0.46	1.55	0.12	181.2
COLY 2013	4.89	21.32	3.87	38.41	84.61	1.73	8.30	4.74	1.04	0.63	0.77	0.62	1.10	0.12	172.2
DOM 2012	2.77	14.34	1.39	17.89	59.03	3.34	10.76	2.29	1.34	0.20	1.20	0.45	1.69	0.03	116.7
DOM 2013	1.79	9.67	1.25	15.70	45.60	2.52	9.01	2.90	0.95	0.27	0.98	0.47	0.89	0.03	92.0
ASMK	2.33	14.84	1.97	15.20	37.40	2.63	8.75	9.25	1.07	0.23	0.19	0.18	0.93	0.12	95.1
BBED	3.49	9.83	1.76	10.71	17.73	2.51	6.27	9.03	1.99	1.60	0.97	0.49	1.77	0.18	68.3
BALPN	1.28	4.07	1.29	8.00	24.18	0.23	1.43	0.85	1.24	0.49	0.43	0.14	0.61	0.03	44.3

BBATH	11.05	10.25	3.94	10.95	55.64	2.29	10.69	4.06	1.26	0.40	0.55	0.42	0.77	1.07	113.3
BEDSF	0.89	4.02	1.00	8.98	23.21	0.49	5.61	4.72	4.32	1.05	0.59	0.49	1.43	0.03	56.8
BMCI	2.30	6.59	1.72	10.87	19.15	2.25	6.20	1.94	2.94	1.61	1.73	0.71	2.05	1.08	61.1
BRAEB	1.61	7.48	1.16	9.45	13.77	1.79	2.37	1.69	1.44	0.51	0.64	0.40	0.92	0.22	43.5
CAMBP	2.62	10.81	1.98	10.90	17.75	1.32	2.70	2.93	1.86	0.97	0.65	0.21	1.03	0.16	55.9
CHRSP	2.38	11.23	1.99	21.12	60.95	1.13	5.61	3.74	7.11	2.49	1.02	0.99	3.40	0.21	123.4
QCOX	1.99	8.18	1.94	11.35	8.16	1.19	2.53	2.48	3.11	0.48	0.64	0.32	1.46	0.59	44.4
COP	2.06	7.85	1.75	9.49	9.76	1.51	2.83	2.13	0.77	0.16	0.29	0.13	0.77	0.16	39.7
DAS 2012	4.19	19.62	2.73	24.93	12.45	2.38	14.78	22.17	1.83	0.70	1.07	0.55	1.99	0.03	109.4
DAS 2013	3.96	17.69	2.88	25.91	13.03	2.10	13.87	21.02	3.29	0.56	1.89	1.22	2.20	0.03	109.7
DAS 2014	2.88	15.74	2.48	19.31	9.71	1.97	12.06	15.09	0.87	0.35	0.64	0.37	0.93	0.03	82.4
DEC	5.45	17.64	3.68	18.62	12.30	1.33	12.38	16.85	2.25	1.22	1.11	0.77	2.56	0.03	96.2
DEVQR 2013	2.60	6.85	1.75	10.51	51.45	0.75	15.74	2.21	1.36	0.24	0.41	0.60	1.11	0.33	95.9
DEVQR 2014	2.77	5.65	1.59	6.88	31.06	1.35	12.06	1.87	1.48	0.37	0.28	0.35	0.91	0.52	67.1
DISC 2012	6.68	11.20	2.38	9.78	41.82	1.15	5.79	2.88	1.23	0.36	0.39	0.27	0.87	0.25	85.1
DISC 2013	4.98	9.46	2.69	11.39	39.49	0.70	7.79	4.19	1.22	0.30	0.41	0.38	0.77	1.01	84.8
DUCHF	3.57	7.51	2.60	10.38	14.70	2.91	9.68	6.57	3.09	0.68	0.91	0.33	1.42	1.12	65.5
EWIND	0.49	8.92	0.53	11.20	11.42	0.53	2.49	1.92	1.80	0.70	0.65	0.43	1.27	0.24	42.6
EGR 2012	2.80	10.00	2.22	14.65	42.30	1.96	8.13	2.63	2.00	0.33	0.49	0.40	1.50	0.03	89.4
EGR 2013	3.87	8.44	2.74	14.32	46.76	2.16	13.46	5.00	3.89	0.68	1.03	1.29	2.66	0.03	106.3
EGR 2014	3.91	13.51	3.53	16.07	49.99	2.73	9.62	3.05	0.82	0.44	0.79	0.51	1.45	0.03	106.5
ELTB	0.95	5.55	0.55	8.07	20.25	0.10	2.86	2.63	2.93	0.36	0.62	0.26	1.24	0.54	46.9
EPIC	0.09	0.16	0.11	0.10	24.21	0.21	6.53	5.87	0.93	0.26	0.64	0.29	0.97	0.03	40.4
FALS	0.87	7.81	0.93	11.96	20.50	0.19	3.88	3.59	2.74	0.35	1.04	0.28	1.08	0.54	55.8
GDEL 2012	1.02	6.73	0.90	8.13	12.61	0.76	6.24	1.62	3.30	0.62	1.42	0.43	1.35	0.03	45.2
GDEL 2013	1.20	6.41	1.01	11.78	16.51	1.00	8.49	4.22	3.51	0.40	2.58	0.96	1.71	0.03	59.8
GALA	1.99	6.88	1.60	9.30	19.13	1.12	4.40	4.58	2.39	0.37	1.15	0.62	1.18	0.60	55.3
GKNOB	2.09	9.26	2.35	10.74	32.18	0.58	8.43	7.60	0.61	0.27	0.13	0.19	0.37	0.03	74.8
GPIP	4.03	5.87	3.88	9.88	18.03	1.78	6.18	5.19	7.38	0.69	0.62	0.40	1.35	0.03	65.3
HISTF 2012	1.52	5.41	1.13	12.43	19.48	4.61	3.87	2.15	2.31	0.92	0.83	0.73	1.14	0.80	57.3

HISTF 2013	1.52	6.55	1.08	14.88	20.39	5.42	4.09	3.07	1.41	0.71	0.68	0.58	0.91	0.03	61.3
JGR	1.40	9.25	1.23	14.80	15.64	0.67	6.47	7.01	5.20	0.74	1.40	0.61	1.56	0.25	66.2
JAZZ	0.96	5.08	0.81	7.44	23.38	0.44	1.65	1.38	2.02	0.39	0.77	0.70	1.04	0.66	46.7
KCODL	2.35	8.77	1.31	10.13	24.95	2.10	8.16	0.90	0.63	0.21	0.76	0.23	0.78	0.03	61.3
LADYL	2.51	5.57	1.47	8.96	19.88	0.14	3.02	1.78	4.37	1.41	0.65	0.38	1.33	0.21	51.7
LAXP	10.22	12.05	4.02	10.49	16.02	1.82	2.93	2.28	4.06	0.83	1.09	0.75	2.59	0.42	69.6
LODGN	2.18	5.70	2.32	7.68	26.05	1.79	3.16	2.21	3.04	0.84	0.38	0.44	1.51	0.15	57.4
LORDL	2.14	4.63	1.17	7.37	16.18	0.14	2.93	1.75	4.15	1.33	0.57	0.38	1.26	0.24	44.2
OPMK	1.95	7.28	2.34	8.91	10.62	1.15	4.50	4.38	1.35	0.22	0.25	0.15	0.74	0.16	44.0
PNOIR	3.67	10.07	2.44	13.97	76.45	1.19	7.70	2.73	3.08	1.08	0.75	0.46	1.52	1.21	126.3
REDSL 2012	3.40	16.17	1.51	12.02	39.70	0.36	8.74	5.56	3.12	0.48	2.39	0.82	3.27	0.54	98.1
REDSL 2013	3.00	11.80	1.92	13.28	47.16	0.26	11.89	8.85	2.70	0.49	1.86	1.13	2.40	0.60	107.4
RIBP 2012	3.49	10.48	2.78	14.37	29.14	1.54	8.83	9.55	1.48	0.30	0.69	0.50	0.96	0.03	84.1
RIBP 2013	1.69	6.66	1.37	7.09	28.10	1.89	6.53	8.02	2.19	0.70	0.78	0.49	1.40	0.03	66.9
RGALA	1.77	8.10	1.33	8.21	15.43	0.96	3.40	2.61	5.79	1.06	1.52	0.74	2.39	1.91	55.2
SCRUMP 2012	0.09	0.93	0.11	0.87	54.64	3.10	8.03	6.36	7.08	1.79	2.78	1.11	4.20	1.89	93.0
SCRUMP 2013	0.09	0.90	0.11	2.55	64.83	2.37	11.42	8.55	2.87	0.50	1.68	0.86	2.11	0.82	99.6
SPRTN	3.29	8.82	1.35	7.07	19.09	0.78	4.13	0.66	0.66	0.14	0.18	0.12	0.55	0.29	47.1
THORLP	4.83	7.40	2.02	6.97	24.13	0.74	5.43	2.97	2.82	0.77	0.67	0.50	1.65	0.13	61.0
WHLR 2012	2.06	13.56	1.80	18.09	57.93	1.88	9.61	3.22	0.15	0.27	0.17	0.14	0.22	0.03	109.1
WHLR 2014	2.09	13.21	2.16	18.90	53.11	0.84	11.42	3.26	0.12	0.14	0.12	0.11	0.12	0.03	105.6
WRP 2012	6.58	12.80	3.24	10.26	25.05	0.46	4.60	0.91	2.63	0.72	0.54	0.76	1.43	0.82	70.8
WRP 2013	6.61	14.56	4.58	22.87	53.27	1.25	10.78	4.69	2.96	0.45	0.80	0.90	2.03	1.02	126.8
MAYP	0.09	3.51	0.11	4.06	54.18	0.87	10.27	6.27	2.16	1.00	1.37	0.53	1.23	6.77	92.4
KRAST	0.77	1.98	0.13	2.68	36.91	1.07	4.37	1.95	1.01	0.20	0.56	0.24	0.54	3.37	55.8
BRGCB	0.09	1.19	0.11	3.50	52.43	0.85	21.87	13.29	8.79	2.28	4.56	4.03	7.60	4.37	125.0
WSLCB	51.96	40.59	8.86	18.84	44.42	8.24	9.50	2.76	2.95	0.85	0.55	0.46	1.78	1.18	192.9
NDER 2012	11.98	34.06	8.04	46.22	29.32	2.63	10.02	10.38	3.69	1.22	1.12	0.93	1.39	2.91	163.9
NDER 2013	16.41	48.79	8.70	50.83	24.79	4.72	10.51	8.16	3.61	1.08	0.73	0.51	1.11	3.88	183.8
NVCP 2012	12.89	47.10	11.90	51.62	62.23	5.41	24.70	19.70	5.07	0.85	1.18	0.97	3.21	4.24	251.1

NVCP 2014	17.85	46.77	12.23	62.72	61.82	6.21	21.40	18.41	4.53	1.38	1.54	1.06	3.12	2.85	261.9
RDFL 2012	4.63	32.61	3.64	68.28	57.24	1.64	8.47	7.01	4.84	1.30	1.24	1.14	1.89	7.06	201.0
RDFL 2013	5.50	34.20	2.81	62.06	55.53	1.81	7.46	4.94	6.27	1.96	0.93	0.78	2.32	7.00	193.5
ROYAL	0.96	13.53	0.99	16.45	422.96	6.07	91.60	9.08	37.25	7.42	21.43	8.76	16.91	107.40	760.8

[a] Data have been backtransformed from log10 transformation. Transformed data including least significant difference (lsd) are available in the supplementary material (**Supporting material S3).** [b] CAT = (+)-catechin, EPIC = (-)-epicatechin, PROC B1 = procyanidin B1, PROC B2 = procyanidin B2, CA = chlorogenic acid, PQCA = *p*-coumaroylquinic acid, PHLOR = phloridzin, PHLXG = phloretin-2'-*O*-xylosylglucoside, QGAL = quercetin galactoside, QGLU = quercetin glucoside, QXYL = quercetin xyloside, QUER = quercitrin, AVIC = avicularin, ΣPC = sum of phenolic compounds. For explanation of cultivar name abbreviations refer to **Table 1**.

Table 3. Phenolic composition of apple peel (mg 100 g^{-1} FW)[a]

Cultivar	CAT	EPIC	PROC B1	PROC B2	CA	CRPT[b]	PHLOR	PHLXG	ΣQGLYC[c]	IDEAIN	ΣPC
CBR	5.54^{d}	30.14^{e}	11.05^{g}	46.49^{fg}	39.07^{j}	1.38^{e}	23.40^{h}	12.14^{i}	150.76^{hi}	ND*	320.0^{fg}
CUMN	8.75^{fg}	35.09^{ef}	9.41^{e}	27.05^{e}	21.60^{hi}	0.79^{c}	15.83^{g}	5.80^{cde}	238.17^{j}	2.26^{b}	364.8^{gh}
PPRF	10.35^{g}	43.65^{gh}	13.62^{g}	49.82^{g}	23.69^{hi}	3.97^{h}	13.95^{fg}	2.04^{a}	125.16^{fghi}	18.45^{de}	304.7^{ef}
STJR	17.91^{h}	113.45^{k}	24.22^{i}	107.32^{j}	144.71^{m}	1.72^{a}	74.78^{kl}	10.40^{i}	118.38^{fgh}	6.90^{c}	619.8^{j}
GMT	5.63^{d}	30.07^{e}	7.50^{f}	41.31^{f}	46.29^{jk}	1.03^{d}	11.09^{de}	5.51^{cd}	144.35^{ghi}	3.10^{b}	295.9^{def}
BRAM	7.07^{ef}	53.47^{i}	7.20^{f}	65.49^{h}	11.79^{fg}	2.47^{g}	8.98^{d}	9.75^{hi}	41.14^{bc}	ND	207.4^{bc}
DAS	5.54^{d}	50.07^{hi}	6.69^{ef}	47.03^{fg}	2.18^{c}	1.72^{a}	82.43^{l}	54.75^{l}	50.45^{bcde}	ND	300.9^{def}
EGR	5.73^{de}	33.51^{ef}	6.23^{ef}	43.78^{fg}	55.94^{k}	1.20^{de}	45.94^{j}	6.76^{fg}	49.13^{bcd}	ND	248.2^{cde}
WHLR	7.28^{f}	56.75^{i}	11.84^{g}	73.74^{hi}	76.59^{l}	1.80^{f}	133.60^{m}	10.46^{i}	9.01^{a}	ND	381.1^{h}
HISTF	3.67^{c}	20.70^{c}	3.34^{d}	28.02^{e}	1.96^{b}	1.07^{d}	3.96^{a}	1.91^{a}	110.88^{fgh}	1.62^{a}	177.1^{ab}
OPMK	5.08^{d}	24.46^{d}	2.69^{cd}	18.99^{d}	0.10^{a}	0.44^{b}	12.25^{ef}	7.10^{efg}	59.10^{de}	2.07^{b}	132.3^{a}
REDSL	3.05^{c}	44.11^{gh}	2.39^{c}	25.57^{e}	26.77^{i}	1.72^{a}	5.54^{b}	4.87^{c}	115.45^{fgh}	13.43^{d}	242.9^{cd}
RGALA	3.28^{c}	37.55^{fg}	2.94^{cd}	28.39^{e}	6.77^{d}	1.72^{a}	5.02^{b}	5.43^{cd}	167.28^{i}	24.21^{efg}	282.6^{def}
SCRUMP	0.09^{a}	7.22^{a}	0.11^{a}	8.22^{b}	19.13^{h}	0.53^{b}	10.09^{de}	6.32^{def}	105.21^{fg}	20.46^{def}	177.4^{ab}
KRAST	1.83^{b}	13.56^{b}	1.31^{b}	15.50^{c}	14.37^{g}	0.68^{c}	7.17^{c}	3.53^{b}	40.10^{b}	34.04^{g}	132.1^{a}
MAYP	0.09^{a}	6.32^{a}	0.11^{a}	0.10^{a}	41.78^{j}	5.16^{ij}	16.27^{g}	7.18^{efg}	57.46^{cde}	24.60^{efg}	159.1^{ab}
NDER	25.39^{i}	160.36^{l}	20.25^{hi}	111.71^{j}	9.78^{ef}	2.53^{g}	60.76^{k}	24.97^{k}	121.93^{fghi}	62.37^{h}	600.1^{j}
NVCP	14.59^{h}	81.87^{j}	17.95^{h}	80.91^{i}	27.01^{i}	5.08^{j}	34.55^{i}	17.01^{j}	70.11^{e}	22.28^{efg}	371.4^{gh}
RDFL	9.67^{g}	72.24^{j}	7.01^{f}	85.02^{i}	58.61^{kl}	5.26^{j}	28.71^{hi}	8.15^{gh}	100.09^{f}	88.92^{h}	463.7^{i}
WSLCB	83.66^{j}	215.63^{m}	23.26^{i}	78.69^{i}	8.83^{de}	4.30^{hi}	85.15^{l}	8.07^{gh}	153.04^{hi}	30.90^{fg}	691.5^{k}

[a] Data have been backtransformed from log10 transformation. Means within the same column with no letters in common are significantly different based on the ANOVA analysis for the log10 transformed data. [b] CRPT = cryptochlorogenic acid, [c] ΣQGLYC = sum of quercetin glycosides. [*] ND = Not Detected. For explanation of other abbreviations refer to **Table 1 and Table 2**.

Table 4. Phenolic composition of apple flesh (mg 100 g^{-1} FW)[a]

Cultivar	CAT	EPIC	PROC B1	PROC B2	CA	PCQA	PHLOR	PHLXG	ΣQGLYC	IDEAIN	ΣPC
CBR	1.96^{d}	14.51^{gh}	1.90^{de}	26.67^{h}	47.21^{f}	3.98^{i}	1.33^{d}	2.05^{ghi}	0.41^{bc}	ND*	100.0^{e}
CUMN	4.73^{g}	13.61^{fgh}	4.78^{g}	14.01^{ef}	30.31^{de}	2.15^{g}	1.16^{cd}	1.01^{de}	0.28^{ab}	ND	72.0^{d}
PPRF	8.29^{h}	23.51^{i}	6.76^{i}	39.45^{i}	46.24^{f}	12.47^{k}	1.10^{cd}	0.50^{b}	tr**	ND	138.3^{fg}
STJR	8.13^{h}	85.07^{m}	7.03^{i}	102.57^{l}	140.60^{i}	0.74^{b}	8.13^{j}	3.44^{k}	0.66^{cde}	ND	356.4^{k}
GMT	4.01^{fg}	11.87^{fg}	2.75^{f}	25.79^{h}	75.96^{h}	0.55^{a}	4.33^{i}	2.40^{ij}	0.68^{cde}	ND	128.3^{f}
BRAM	7.76^{h}	11.24^{f}	4.63^{h}	14.59^{f}	29.38^{d}	2.26^{gh}	1.28^{d}	0.92^{cde}	tr	ND	72.1^{d}
DAS	3.60^{f}	15.27^{h}	2.84^{f}	25.11^{h}	4.29^{a}	1.69^{f}	3.10^{fghi}	12.13^{m}	0.29^{ab}	ND	68.3^{cd}
EGR	3.47^{f}	9.01^{de}	2.55^{f}	14.51^{f}	41.44^{ef}	2.23^{g}	2.88^{fgh}	1.14^{e}	0.24^{a}	ND	77.5^{d}
WHLR	2.44^{de}	11.56^{f}	2.06^{e}	19.91^{g}	60.19^{g}	1.09^{cd}	0.97^{bcd}	1.56^{fg}	tr	ND	99.8^{e}
HISTF	1.45^{c}	4.67^{c}	1.11^{b}	12.37^{ef}	16.88^{c}	5.58^{j}	0.55^{a}	0.35^{a}	tr	ND	43.0^{ab}
OPMK	2.12^{de}	7.45^{d}	1.90^{de}	11.75^{e}	4.72^{a}	1.05^{c}	0.71^{ab}	1.71^{gh}	tr	ND	31.4^{a}
REDSL	2.64^{e}	11.15^{ef}	1.40^{c}	9.26^{d}	30.99^{d}	1.42^{ef}	0.67^{ab}	1.20^{ef}	tr	ND	58.7^{bcd}
RGALA	1.98^{d}	4.85^{c}	1.64^{cd}	6.13^{c}	13.13^{b}	0.80^{b}	0.85^{bc}	0.81^{cd}	tr	ND	30.2^{a}
SCRUMP	0.09^{a}	0.16^{a}	0.11^{a}	0.10^{a}	35.74^{de}	2.79^{h}	2.18^{ef}	2.17^{hij}	0.35^{bc}	ND	43.70^{ab}
KRAST	0.09^{a}	1.19^{b}	0.11^{a}	2.30^{b}	35.70^{de}	1.33^{de}	0.76^{ab}	0.74^{c}	0.58^{cde}	0.47^{a}	43.3^{ab}
MAYP	0.63^{b}	0.16^{a}	0.11^{a}	0.10^{a}	42.21^{ef}	1.35^{de}	1.98^{e}	2.88^{jk}	0.88^{def}	0.68^{a}	51.0^{abc}
NDER	17.53^{i}	43.17^{kl}	9.31^{j}	48.52^{j}	18.27^{c}	5.16^{j}	3.52^{ghi}	5.47^{l}	1.09^{ef}	2.07^{b}	154.1^{gh}
NVCP	19.01^{i}	32.77^{j}	10.57^{j}	45.19^{ij}	41.49^{ef}	6.11^{j}	2.66^{efg}	6.98^{l}	tr	ND	164.8^{hi}
RDFL	4.64^{g}	35.55^{jk}	3.57^{g}	71.73^{k}	61.92^{gh}	2.38^{gh}	1.94^{e}	3.53^{k}	1.48^{f}	3.864^{c}	190.6^{i}
WSLCB	91.81^{j}	50.44^{l}	20.35^{k}	25.23^{h}	71.25^{gh}	14.80^{k}	3.94^{hi}	2.69^{ijk}	0.47^{bcd}	0.78^{a}	281.8^{j}

[a] Data have been backtransformed from log10 transformation. Means within the same column with no letters in common are significantly different based on the ANOVA analysis for the log10 transformed data. * ND = Not Detected, ** tr = traces. For explanation of cultivar name and phenolic compound abbreviations refer to **Table 1 and Table 2.**

Table 5. Phenolic composition of apple seeds (mg 100 g^{-1} FW)[a,b]

Cultivar	CAT	EPIC	PROC B1	PROC B2	CA	PHLOR	PHLXG	IDEAIN	ΣPC	AMYG[c]	MASS/ FRUIT	SEEDS/ FRUIT
CBR	0.69^{a}	21.37^{fgh}	4.05^{a}	40.41^{i}	119.1^{gh}	955.0^{e}	87.5^{jk}	ND	1228.1^{ef}	194.00^{g}	0.31	5
CUMN	tr*	6.18^{abc}	tr	25.72^{def}	29.9^{c}	538.3^{cd}	28.6^{f}	ND	628.7^{d}	136.30^{defg}	0.53	12
PPRF	tr	7.60^{abcde}	tr	27.43^{defg}	18.8^{b}	290.4^{a}	14.4^{de}	ND	358.6^{a}	143.50^{efg}	0.36	6
STJR	1.68^{b}	19.77^{fg}	5.20^{ab}	35.94^{hi}	55.60^{e}	477.50^{bc}	4.40^{a}	ND	600.1^{cd}	193.80^{g}	0.53	9
GMT	1.18^{ab}	17.49^{f}	11.47^{b}	39.82^{i}	120.5^{gh}	2113.5^{g}	52.1^{hi}	ND	2356.1^{g}	82.00^{abcde}	0.12	3
BRAM	1.37^{ab}	44.69^{i}	12.21^{b}	56.14^{j}	87.5^{fg}	929.0^{e}	109.4^{k}	ND	1240.3^{ef}	70.00^{abc}	0.07	2
DAS	tr	4.95^{ab}	tr	13.16^{ab}	24.6^{bc}	853.1^{e}	62.2^{ij}	ND	958.0^{e}	80.50^{abcd}	0.33	6
EGR	tr	7.14^{abcd}	tr	26.84^{defg}	70.3^{ef}	408.3^{abc}	29.5^{fg}	ND	542.1^{bcd}	59.90^{ab}	0.31	4
WHLR	0.81^{ab}	26.17^{h}	4.74^{a}	32.63^{gh}	148.6^{hi}	918.3^{e}	6.5^{b}	ND	1137.8^{e}	106.10^{bcdef}	0.09	5
HISTF	tr	17.27^{f}	tr	27.9^{efg}	51.5^{e}	406.4^{abc}	44.8^{hi}	ND	547.9^{bcd}	106.20^{bcdef}	0.49	7
OPMK	tr	12.39^{e}	tr	21.73^{cd}	61.8^{ef}	939.7^{e}	52.8^{hi}	ND	1088.4^{e}	107.70^{bcdef}	0.53	8
REDSL	1.59^{b}	46.49^{i}	11.92^{b}	52.96^{j}	325.1^{j}	751.6^{de}	10.5^{cd}	ND	1200.2^{e}	157.60^{fg}	0.44	8
RGALA	tr	7.73^{abcde}	tr	16.32^{bc}	48.2^{de}	341.2^{ab}	43.1^{ghi}	ND	456.6^{abc}	90.70^{bcde}	0.18	3
SCRUMP	tr	10.75^{cde}	tr	28.87^{fg}	56.0^{e}	451.9^{bc}	43.2^{h}	ND	590.7^{cd}	99.60^{bcdef}	0.31	5
MAYP	tr	8.84^{bcde}	tr	22.41^{cde}	51.2^{e}	400.9^{abc}	7.7^{bc}	12.1^{a}	503.2^{bcd}	120.20^{cdef}	0.18	3
NDER	tr	23.00^{gh}	tr	30.11^{fgh}	32.2^{cd}	305.5^{a}	15.4^{e}	ND	406.2^{ab}	93.10^{bcde}	0.16	3
NVCP	tr	2.84^{a}	tr	7.74^{a}	7.8^{a}	485.3^{bc}	14.8^{de}	26.53^{b}	545.0^{bcd}	104.60^{bcdef}	0.06	3
RDFL	tr	11.14^{de}	tr	21.5^{cd}	59.2^{ef}	373.3^{ab}	79.4^{jk}	23.52^{b}	568.1^{bcd}	138.60^{defg}	0.14	3
WSLCB	2.33^{b}	23.55^{gh}	12.71^{b}	40.17^{i}	221.8^{ij}	1406.0^{fg}	9.1^{bc}	ND	1715.7^{fg}	19.00^{a}	0.14	4

[a] Data have been backtransformed from log10 transformation. Means within the same column with no letters in common are significantly different based on the ANOVA analysis for the log10 transformed data. [b] Most 'Krasnyi Shtandart' apples contained no seeds and therefore this cultivar was not included in the table. [c] AMYG = amygdalin, [*] tr = traces. For explanation of cultivar name and phenolic compound abbreviations refer to **Table 1 and Table 2**.

Table 6. Concentration of soluble sugars in whole apples (g 100 g^{-1} FW)

Cultivar	Fructose	Glucose	Sucrose	Sorbitol[a]		Total sugars
CBR 2012	3.75	1.45	1.77	0.35	(-0.45)	7.3
CBR 2013	5.45	1.85	3.62	1.77	(0.25)	12.7
CUMN 2012	5.35	2.60	0.60	0.59	(-0.23)	9.1
CUMN 2014	5.65	2.70	0.68	0.41	(-0.39)	9.4
PPRF 2012	3.19	0.72	2.90	0.50	(-0.30)	7.3
PPRF 2013	2.42	0.74	1.49	0.33	(-0.48)	5.0
PPRF 2014	4.46	2.17	2.18	0.31	(-0.51)	9.1
DYMR 2012	4.51	2.21	0.92	0.35	(-0.45)	8.0
DYMR 2014	4.83	2.36	1.83	0.22	(-0.65)	9.2
PENB	4.65	1.85	1.11	0.55	(-0.26)	8.2
STJR	1.15	0.40	0.31	0.32	(-0.49)	2.2
GMT 2012	3.21	0.64	3.07	0.50	(-0.30)	7.4
GMT 2013	4.37	1.38	4.28	0.37	(-0.43)	10.4
SIW79	4.58	1.38	2.62	0.39	(-0.40)	9.0
SIW92	4.70	1.31	2.12	0.37	(-0.44)	8.5
GSPR	3.96	1.33	1.57	0.12	(-0.93)	7.0
BMOR	2.72	0.43	0.99	0.25	(-0.61)	4.4
BRAM 2012	4.24	0.61	2.74	0.36	(-0.44)	8.0
BRAM 2013	4.02	0.93	2.93	0.24	(-0.61)	8.1
CHARL	4.92	0.49	3.45	0.51	(-0.29)	9.4
COLY 2012	4.83	1.96	2.65	0.73	(-0.14)	10.2
COLY 2013	5.63	2.48	3.67	0.59	(-0.23)	12.4
DOM 2012	2.16	0.48	1.11	0.28	(-0.55)	4.0
DOM 2013	2.41	0.44	2.02	0.12	(-0.93)	5.0
ASMK	3.99	0.93	6.06	0.66	(-0.18)	11.6
BBED	4.12	1.06	2.98	0.38	(-0.43)	8.5
BALPN	5.27	0.77	3.59	0.51	(-0.29)	10.1
BBATH	3.90	0.74	2.92	0.28	(-0.56)	7.8
BEDSF	4.83	1.29	3.67	0.51	(-0.29)	10.3
BMCI	6.38	0.78	2.38	0.49	(-0.31)	10.0
BRAEB	3.72	0.86	2.34	0.43	(-0.37)	7.3
CAMBP	4.78	1.19	1.94	0.57	(-0.24)	8.5
CHRSP	4.37	1.41	3.78	0.56	(-0.25)	10.1
QCOX	4.40	0.87	4.95	0.60	(-0.22)	10.8
COP	4.64	0.75	5.40	0.51	(-0.30)	11.3
DAS 2012	4.56	0.70	3.94	0.65	(-0.19)	9.9
DAS 2013	6.28	0.81	5.62	0.67	(-0.17)	13.4
DAS 2014	5.87	0.92	6.11	0.60	(-0.22)	13.5
DEC	4.12	1.88	1.90	0.62	(-0.21)	8.5
DEVQR 2013	5.22	1.33	1.57	0.89	(-0.05)	9.0
DEVQR 2014	5.11	1.82	1.52	0.23	(-0.64)	8.7
DISC 2012	3.51	0.80	1.77	0.42	(-0.37)	6.5
DISC 2013	5.47	1.65	2.00	0.20	(-0.71)	9.3
DUCHF	3.98	0.58	3.62	0.34	(-0.47)	8.5

EWIND	5.02	0.65	4.38	0.57	(-0.25)	10.6
EGR 2012	5.66	1.03	4.07	0.57	(-0.24)	11.3
EGR 2013	2.96	0.62	1.64	0.38	(-0.42)	5.6
EGR 2014	5.64	1.25	8.72	0.37	(-0.43)	16.0
ELTB	4.12	0.57	1.55	0.65	(-0.18)	6.9
EPIC	3.13	0.45	1.69	0.32	(-0.50)	5.6
FALS	5.00	0.83	4.19	0.46	(-0.34)	10.5
GDEL 2012	5.57	1.24	3.69	0.64	(-0.19)	11.1
GDEL 2013	6.96	1.81	2.34	0.23	(-0.65)	11.3
GALA	5.17	0.80	2.52	0.23	(-0.64)	8.7
GKNOB	4.46	0.56	4.37	0.58	(-0.24)	10.0
GPIP	4.64	0.72	3.91	0.57	(-0.24)	9.8
HISTF 2012	5.49	0.92	2.66	0.46	(-0.33)	9.5
HISTF 2013	3.99	0.59	2.17	0.19	(-0.73)	6.9
JGR	5.23	0.87	3.15	0.22	(-0.65)	9.5
JAZZ	4.38	0.76	4.34	0.42	(-0.38)	9.9
KCODL	2.76	0.37	2.28	0.22	(-0.66)	5.6
LADYL	4.76	0.95	1.73	0.49	(-0.31)	7.9
LAXP	4.97	0.69	2.50	0.46	(-0.34)	8.6
LODGN	5.41	0.76	2.89	0.68	(-0.17)	9.7
LORDL	5.03	1.44	1.97	0.41	(-0.39)	8.9
OPMK	4.52	1.21	2.50	0.37	(-0.43)	8.6
PNOIR	3.63	1.84	1.12	0.37	(-0.43)	7.0
REDSL 2012	4.43	0.51	2.88	0.36	(-0.45)	8.2
REDSL 2013	4.68	1.10	1.48	0.81	(-0.09)	8.1
RIBP 2012	3.10	0.63	3.22	0.37	(-0.32)	7.3
RIBP 2013	4.36	0.94	4.35	0.48	(-0.43)	10.1
RGALA	5.47	0.98	3.99	0.54	(-0.27)	11.0
SCRUMP 2012	5.79	2.32	1.82	0.34	(-0.46)	10.3
SCRUMP 2013	5.85	1.70	2.31	0.26	(-0.59)	10.1
SPRTN	5.28	0.96	3.05	0.46	(-0.33)	9.8
THORLP	3.01	0.42	2.05	0.52	(-0.28)	6.0
WHLR 2012	5.29	1.03	6.08	0.81	(-0.28)	13.2
WHLR 2014	3.72	0.71	2.66	0.52	(-0.09)	7.6
WRP 2012	5.54	1.14	2.73	0.47	(-0.33)	9.9
WRP 2013	5.72	1.76	1.80	0.23	(-0.65)	9.5
MAYP	4.67	1.65	2.05	0.73	(-0.13)	9.1
KRAST	3.70	0.73	2.51	0.41	(-0.39)	7.3
BRGCB	3.90	1.22	6.28	0.75	(-0.12)	12.2
WSLCB	2.13	0.42	1.12	0.28	(-0.56)	3.9
NDER 2012	2.33	0.39	1.15	0.22	(-0.32)	4.1
NDER 2013	3.75	0.71	1.93	0.48	(-0.65)	6.9
NVCP 2012	4.56	1.09	2.91	0.56	(-0.25)	9.1
NVCP 2014	5.63	1.31	3.14	0.55	(-0.26)	10.6
RDFL 2012	2.97	1.03	3.72	0.75	(-0.28)	8.5
RDFL 2013	2.24	0.95	2.81	0.52	(-0.13)	6.5

ROYAL	2.75	1.85	6.08	1.99	(0.30)	12.7
lsd[b]	0.86	0.30	0.74		(0.099)	1.6

[a] Data have been backtransformed from log10 transformation, transformed data and significant difference (lsd) have been included in brackets. [b] lsd = least significant difference. For explanation of cultivar name abbreviations refer to **Table 1**.

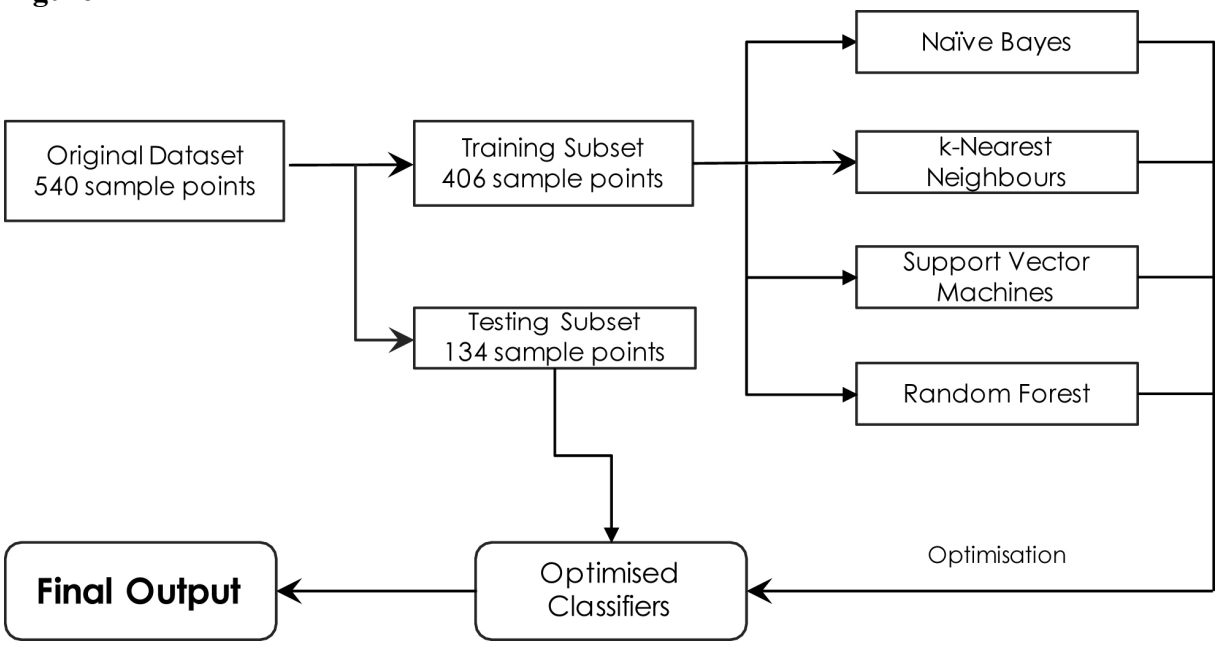
Table 7. Organic acid composition (mg 100 g^{-1} FW) of 20 apple cultivars

Cultivar	Malic acid	Quinic acid	Ascorbic acid	ΣOA[a]
CBR	1249.9	107.6	ND	1357.5
CUMN	152.9	134.1	ND	287.0
STJR	150.4	168.6	ND	319.0
PPRF	628.9	90.2	2.19	721.3
GMT	2277.1	144.7	5.53	2427.3
BRAM	1978.9	117.1	3.4	2099.4
DAS	533.3	83.6	ND	616.9
EGR	911.4	85.8	ND	997.2
WHLR	789.3	18.2	ND	807.5
HISTF	1020.2	127.4	1.18	1148.8
OPMK	620.8	54.9	0.36	676.1
REDSL	341.2	131.1	ND	472.3
RGALA	333.0	38.9	1.47	373.4
SCRUMP	542.5	107.6	ND	650.1
KRAST	1716.9	109.4	2.63	1828.9
MAYP	2709.8	148.9	ND	2858.7
NDER	2154.1	209.5	ND	2363.6
NVCP	1008.9	113.1	1.15	1123.2
RDFL	650.1	98.5	4.91	753.5
WSLCB	408.2	173.3	ND	581.5
lsd[b]	87.34	15.15	2.102	102.5

[a] ΣOA = sum of organic acids, [b] lsd = least significant difference. For explanation of cultivar name abbreviations refer to **Table 1**.

FIGURE GRAPHICS

Figure 1



1

Figure 2

2

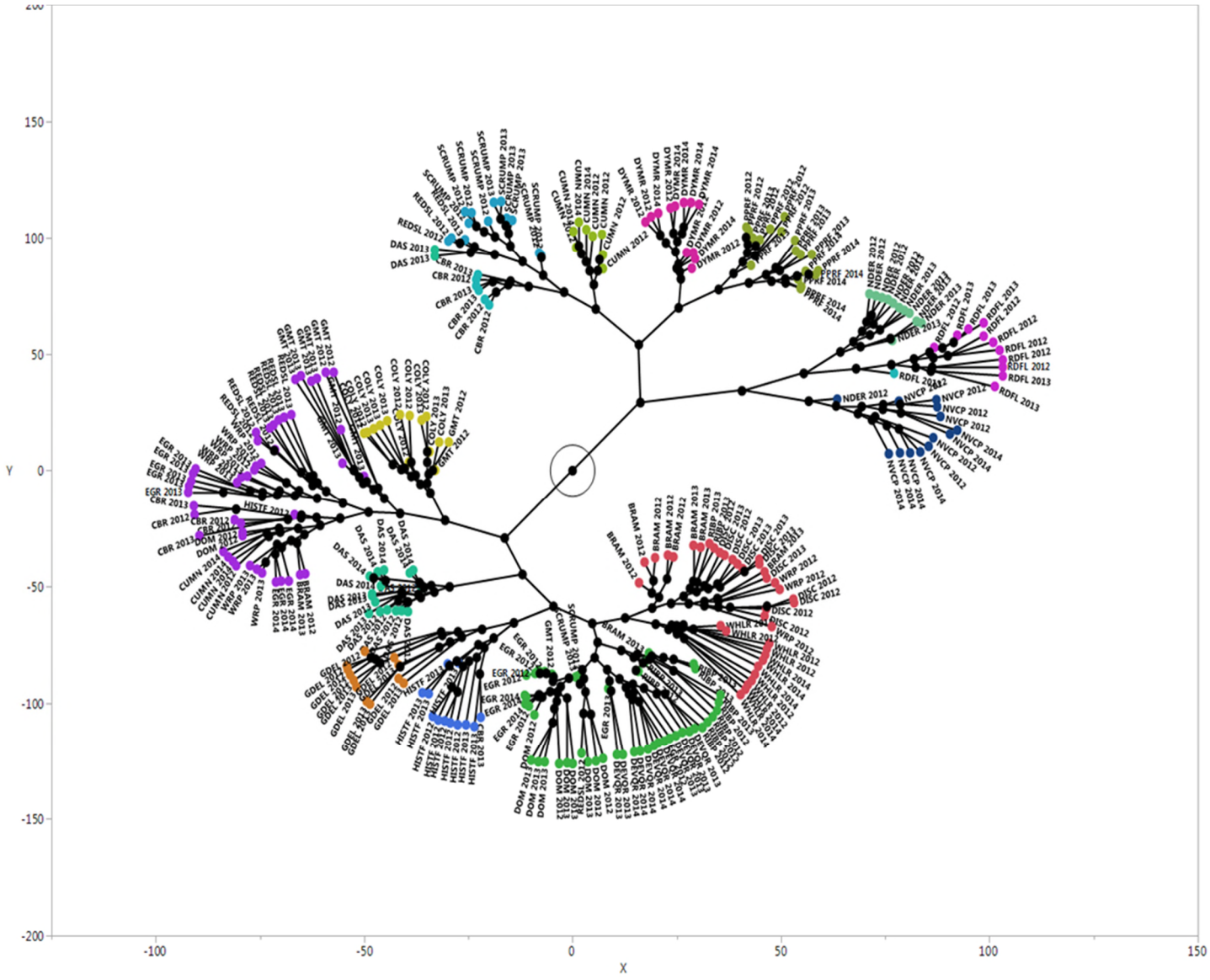


Figure 3

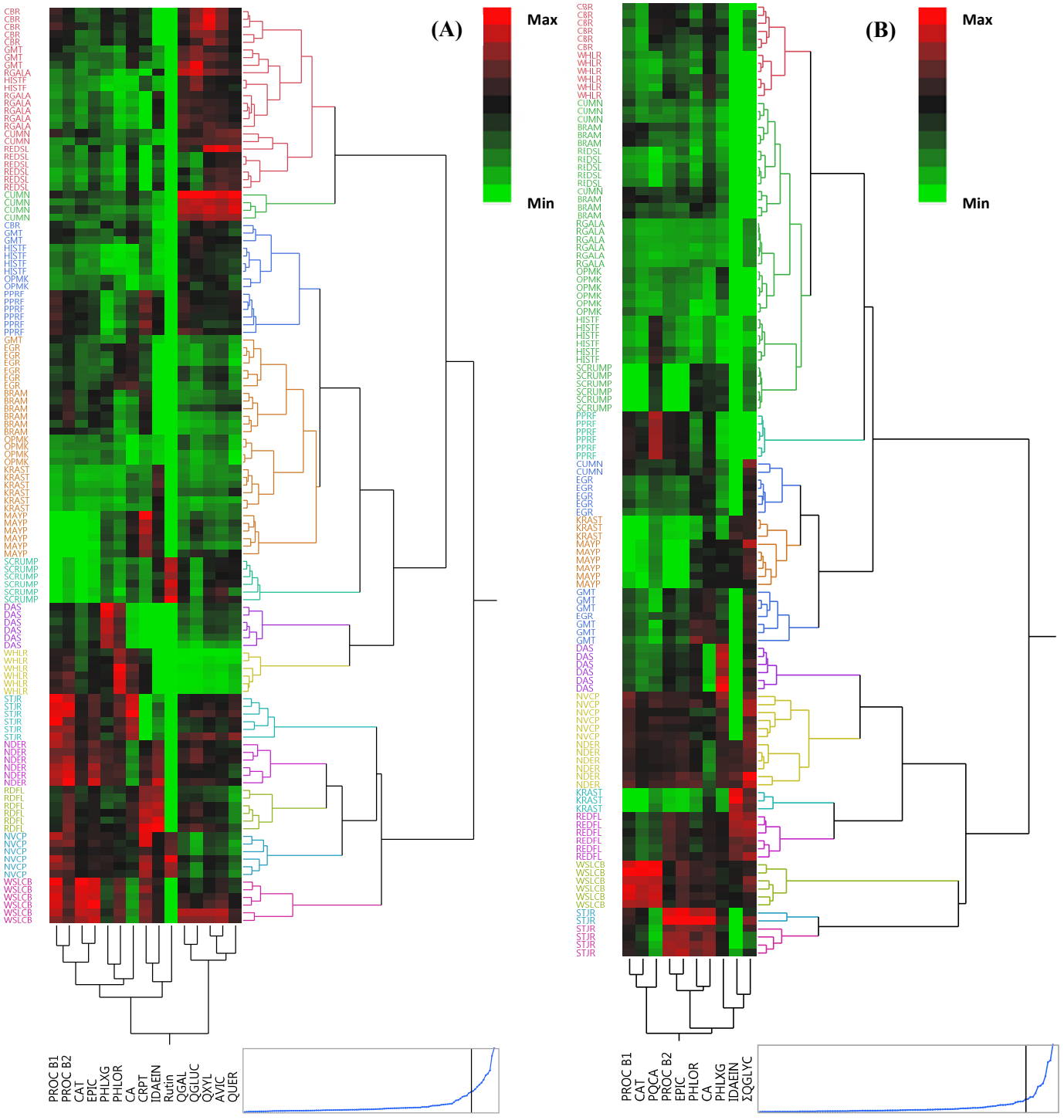

(A)
Max
Min
PROC B1
PROC B2
CAT
EPIC
PHLXG
PHLOR
CA
CRPT
IDAEIN
Rutin
QGAL
QGLUC
QXYL
AVIC
QUER
(B)
Max
Min
PROC B1
CAT
POCA
PROC B2
EPIC
PHLOR
CA
PHLXG
IDAEIN
ΣQGLYC

Figure 4

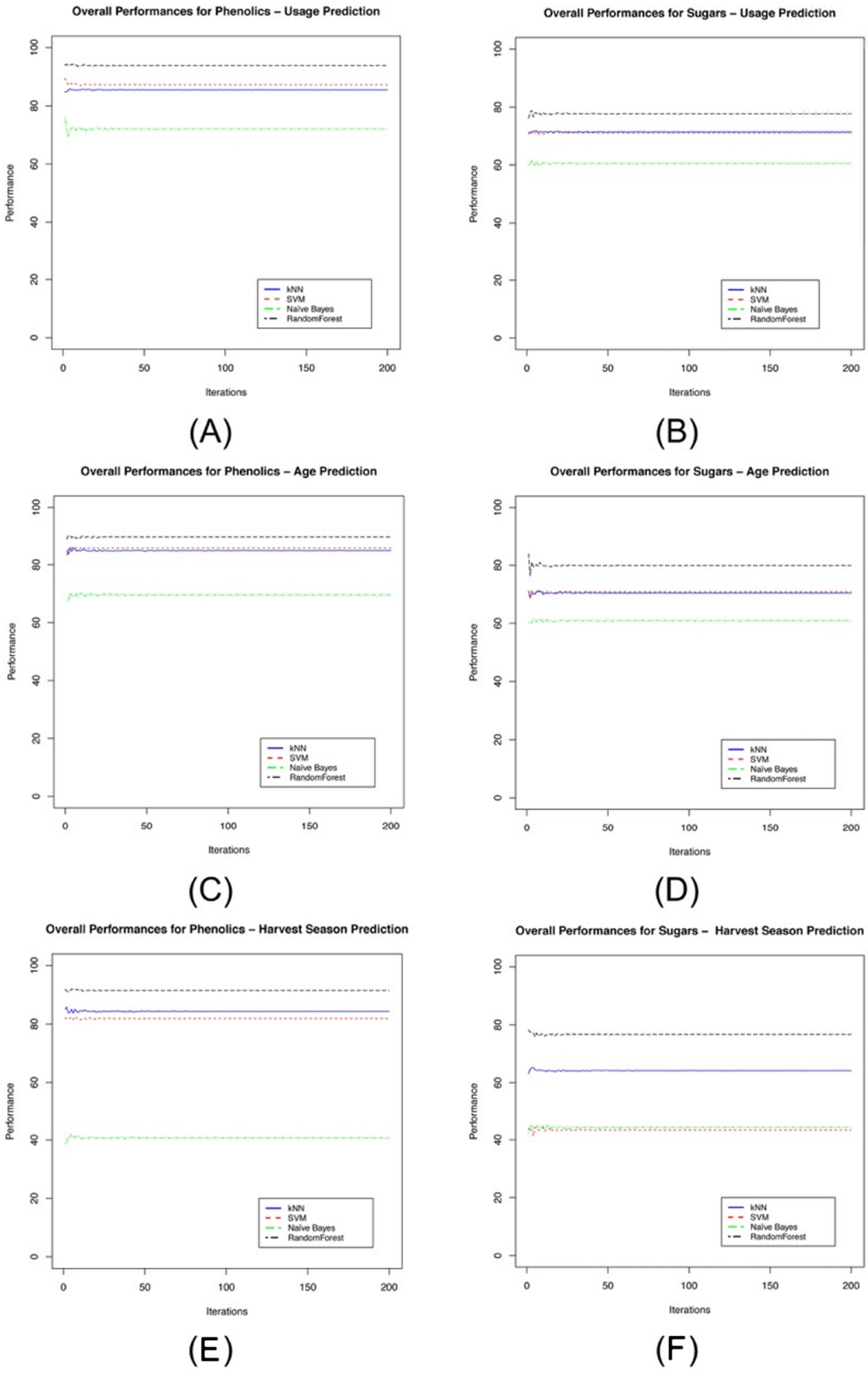
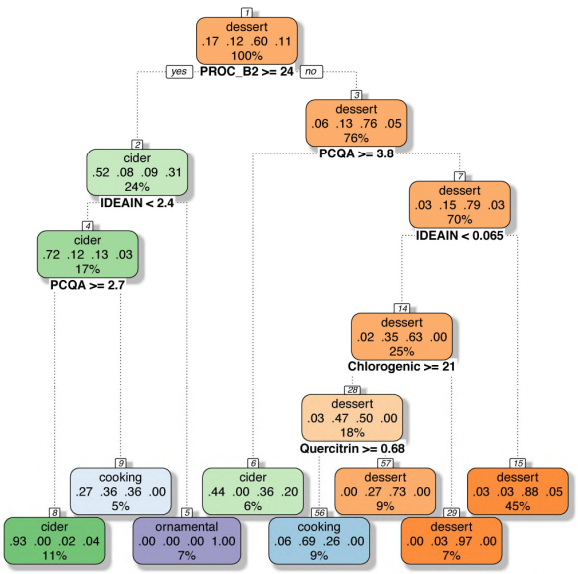
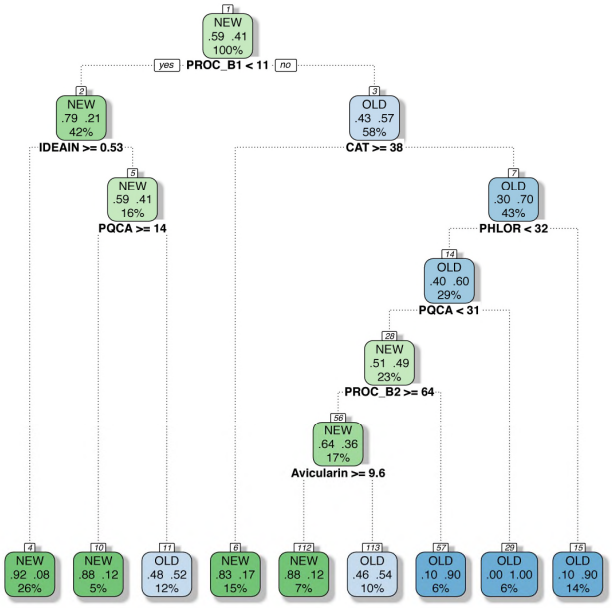

Overall Performances for Phenolics – Usage Prediction
Performance
Iterations
kNN
SVM
Naïve Bayes
RandomForest
(A)
Overall Performances for Sugars – Usage Prediction
Performance
Iterations
kNN
SVM
Naïve Bayes
RandomForest
(B)
Overall Performances for Phenolics – Age Prediction
Performance
Iterations
kNN
SVM
Naïve Bayes
RandomForest
(C)
Overall Performances for Sugars – Age Prediction
Performance
Iterations
kNN
SVM
Naïve Bayes
RandomForest
(D)
Overall Performances for Phenolics – Harvest Season Prediction
Performance
Iterations
kNN
SVM
Naïve Bayes
RandomForest
(E)
Overall Performances for Sugars – Harvest Season Prediction
Performance
Iterations
kNN
SVM
Naïve Bayes
RandomForest
(F)

Figure 5



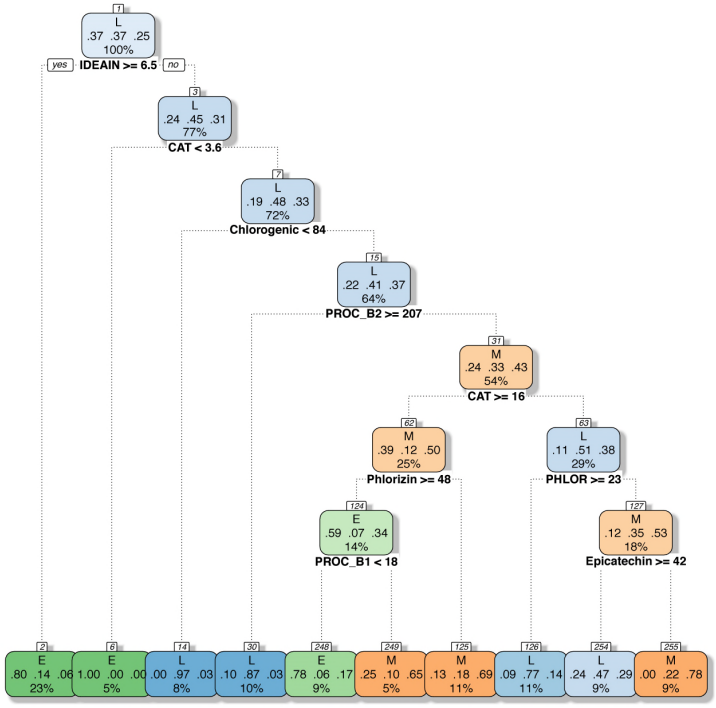
Decision Tree for Phenolic Model for Usage Prediction – (minsplit = 50)

(A)



Decision Tree for Phenolic Model for Age Prediction – (minsplit = 50)

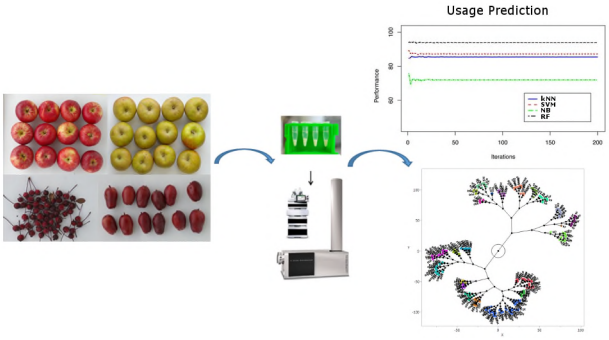
(B)



Decision Tree for Phenolic Model for Harvest Season Prediction – (minsplit = 50)

(C)

FOR TABLE OF CONTENTS ONLY



3