CHAPTER 2

LITERATURE REVIEW

2.1 Background

Dates fruit is scientifically known as *Phoenix dactylifera L.* which belongs to the palm (Arecaceae) family. According to Baliga et al. (2011), phytochemicals rich in dates flesh including phenolics, sterols, carotenoids, anthocyanins, procyanidins and flavonoids. Furthermore, Kursinszki et al (2011) stated that the ripe dates fruit consist of carbohydrates with high energy content, as well as large amounts of dietary fiber. By consuming dates fruit to break fasting will help to avoid overeating of food. When the body begins to absorb the high nutritional value of the dates flesh, feelings of hunger are pacified. Also, the nervous system can get beneficial from consuming dates flesh, since they have such significant amounts of potassium (El-Sohaimy & Hafez, 2010).

2.2 Stages in Development of Dates Fruits

Generally, the maturity process of dates fruits will pass through five stages of ripening in 7 months. Throughout that period, the colour of the fruit will gradually change from green to yellow and then to reddish brown. According to Arabic practice, the five ripening stages of dates fruit are known as Hababouk, Khimri, Khalal, Rutab and Tamar. The maturity and ripeness of dates fruits commonly will affect its physical and chemical properties such as sweetness, texture, color and chemical composition (Al-Saif et al., 2017). Figure 2.1 shows the dates fruit stage of development.

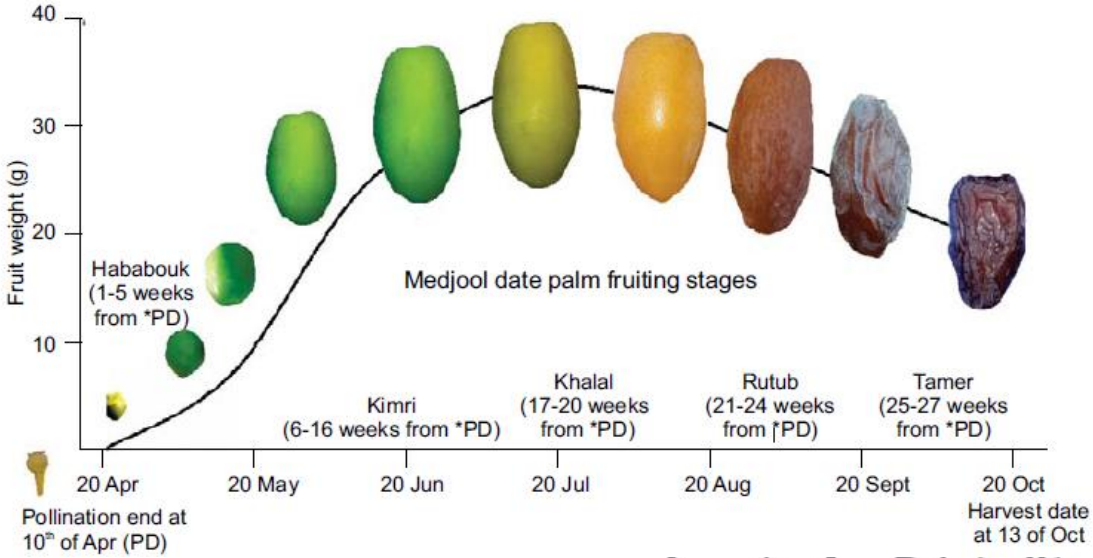


Figure 2.1: Dates Fruit Stage of Development (Al-Hajjaj & Ayyad, 2019)

2.2.1 The Hababouk Stage (First Stage)

The first stage of ripening in the development of dates fruits is known as Hababouk. The stage begins after the fruit set and lasts for four to five weeks (Mortazavi et al., 2015). The immature fruit in this stage is pea sized and weighs about a gram (Baliga et al., 2011). Only few studies have reported on this stage. Figure 2.2 shows the size of fruits accordingly to the stage ripening.

Figure 2.2: Size of Fruits According to the Stage of Ripening (Mortazavi et al., 2015)

2.2.2 The Khimri Stage (Green Stage)

The development of dates fruit in this stage takes nine to fourteen weeks which is known to be the longest stage (Baliga et al., 2011). There is an obvious conversion from a small berry size (hababouk stage) to oblong shape. The color of the fruit is green, the fruit texture is hard, typically bitter and not suitable to consume.

2.2.3 Khalal Stage (Color Stage)

The third stage, Khalal stage which is known as color stage is occur for six weeks in which the fruit is physiologically mature and considered ripe (Baliga et al., 2011). Khalal stage also known as the first three maturity stage of dates fruits where the fruits can be consumed start from this stage (Hussein et al., 2015). A rapid increase of the concentration of non-reducing sugar (mainly sucrose) in dates fruit cause the steady loss of the water content (Diboun et al., 2015). According to Diboun et al (2015), color transition from green color (Khimri) to yellow (Khalal) caused by the degradation of chlorophyll while the yellow color is the marks of carotenoids composition (Mortazavi et al., 2015).

2.2.4 Rutab stage (Soft Ripe Stage)

This stage normally takes place between two to four weeks (Baliga et al., 2011). Rutab stage also called as a soft ripe stage. The fruit texture becomes sweet, soft and delicious (Mortazavi et al., 2015). The sweetness taste of the fruit is elevated as the rate of conversion reducing sugars is greatly increased to form the dominant sugar (Awad et al., 2011). Due to the consistent loss of moisture content, the weight of the fruit decreases as well (Baliga et al., 2011). Some people usually misclassified Rutab stage as Khalal due to the changes of skin color from the top of the fruits to the

down at the connection point to the tree, where the side of the vision may differ (Pourdarbani et al., 2015).

2.2.5 Tamar stage (Full Ripe Stage)

Tamar is the last stage of ripening and physically, the fruit appears dry. The semi-dry and dry dates will have nearly 50 % each of sucrose and reducing sugars. Typically, the flesh is slightly juicy and the skin is wrinkles as shown in Figure 2.2. The color of the skin and of the underlying flesh darkens with time (Baliga et al., 2011). Tamar stage is the best stage of ripening to import or export manufacturing because of the sugar content in dates fruit act as preservative.

2.3 Varieties of Dates Fruits

2.3.1 Ajwa

Ajwa dates are one of the most expensive dates in its raw form in Saudi Arabia. In Tirmidhi (2068) it is quoted that:"Ajwa dates are from paradise"'A'isha reported Allah's Messenger (may peace be upon him) as saying: The Ajwa dates of 'Aliya' contain heating effects and these are antidote in the early morning (Sahih Muslim Book 23, Hadith 5083).

2.3.2 Khalas

Khalas is the most famous dates and is considered one of the finest varieties of dates, Cultivated in the eastern region and is highly consumed and demanded in the market. Khalas has a caramel flavour when cured; delicious when eaten fresh with good taste. This is the variety most commonly taken with Arabic coffee.

2.3.3 Mabroom

These are long red to brownish red dates, deliciously sweet and very chewy. Its price can increase slightly due to the market high demand. Each Mabroom palm tree produces between 150 to 250 kilograms of dates per year. It was originating from Madinah, Saudi Arabia, the Mabroom dates is another dates type which sells in great quantities during Ramadan. Mabroom dates are dry and parched in terms of appearance but they have an exceptionally sweet taste. A bit elastic and chewy in nature, Mabroom dates have a tender interior which is quite rich in taste. It is packaged in the standard 1 kg plastic bags although the sizes can differ with varying demands.

2.3.4 Safawi

Safawi is a soft, cultivated in Madinah and Yanbuae which are also famous for dates; t is dark brown and deliciously sweet and Medina is known for the best and most expensive dates at all, with high demand. Safawi dates are a mine of vitamins, and they are called mine due to their large content of metal elements such as phosphorus, calcium, iron, magnesium, sodium, sulphur, and chlorine.

2.3.5 Mariami

Mariami is one of the varieties which grows in south of Iran in the north area. Palms are grown organically with no chemical treatment during the growing season. The washing stage is completely performed. Fructose is the sugar found mainly in the dates which undergoes metabolic processing in the liver and is known as one of the best sweeteners for diabetics to use. This dates is semi dried with moisture under 15

%. It is dark brown in color and round in size. Shelf life at room temperature is around 18 months, under interim fumigation. Harvest Time is around mid-September.

2.4 Bioactive Compounds in Dates Fruit

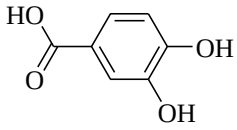
2.4.1 Phenolics

Baliga et al. (2011) stated that phenolic acids constitute one of the main classes of secondary metabolites and in recent years have been a subject of intense study. Phenolic acids contain a hydroxylated benzene ring with one or more carboxyl groups directly or indirectly attached to it. Phenolic acids in Algerian dates are *p*-coumaric acid, ferulic acid and sinapic acid and cinnamic acid and its two derivatives; 5-*o*-caffeoylshikimic acid and xantoxylin (Mansouri et al., 2005). Meanwhile, Al-Farsi et al (2005) stated that 3 varieties of Omani dates have shown the presence of free and bound phenolic acid.

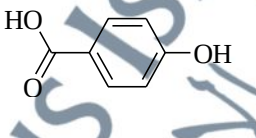
The free phenolic acids such as protocatechuic acid, vanilic acid, syringic acid and ferulic acid and the examples of bound phenolic acid such as gallic acid, caffeic acid, p-coumaric acid and o-coumaric acid. Al Farsi et al. (2005) and Benmeddour et al (2013) who studied Omani and Algerian dates, respectively, stated that the results obtained for phenolic content were much higher compared to those reported by Mansouri et al. (2005) who found that total phenolic content of methanolic extracts is 2.49 mg /100 g to 8.36 mg /100 g fresh weight sample. The observed differences may mainly be attributed to the cultivars and extraction conditions such as solvent and ratio material/solvent (Benmeddour et al., 2013) or other various factors such as growing condition, maturity, season, and soil type (Al-Farsi et al., 2005).

On the other hand, Biglari et al. (2008) who studied Iran dates reported that their results were similar level of phenolic content with those Algerian dates palm by

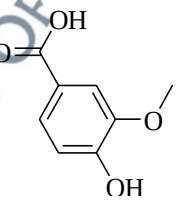
Mansouri et al. (2005). But, one of the types of dates palm fruits studied by Biglari et al. (2008) was several folds higher than other dates; however, level of phenolic content fresh weight of Omani studied by Al-Farsi et al. (2005) was closer to total phenolic content of that type of dates palm fruits. Level of total phenolic content of fresh and dried dates fruits are high (Wu et al., 2004) but Al-Farsi et al. (2005) stated that total phenolic content in dried is much higher than in fresh. In addition, total phenolic content in dates fruits is much higher compared to both fresh fruits and dried fruits (Benmeddour et al., 2013; Kchaou et al., 2013). The examples of the phenolic acids were shown in Figure 2.3.



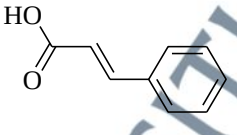
protocatchnic acid



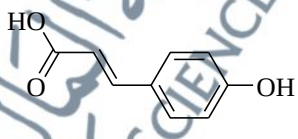
p-hydrobenzoic acid



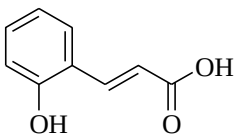
vanillic acid



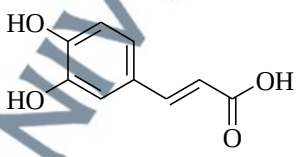
cinnamic acid



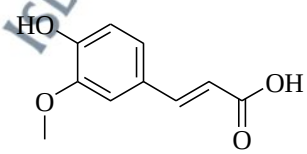
p-coumaric acid



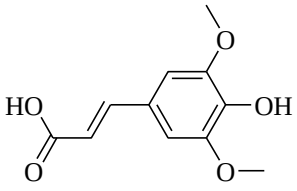
o-coumaric acid



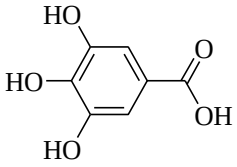
cafeic acid



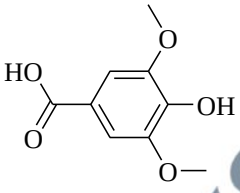
ferulic acid



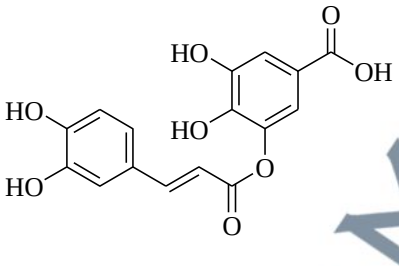
sinapic acid



gallic acid



syringic acid



5-*o*-caffeoylshikimic acid

Figure 2.3: Structure of Phenolic Acids Present in Dates (Baliga et al., 2011)

2.4.2 Anthocyanins

Anthocyanin is one of antioxidant compounds in dates fruits (Samad et al., 2016). Figure 2.4 shows the structure of anthocyanins in dates. Anthocyanins may appear in red, purple, violet or blue (Baliga et al., 2011; Al-Farsi et al., 2005) and are water soluble vacuolar pigments (Baliga et al., 2011). Anthocyanins are of potential health benefits and widely distributed in many fruits, vegetables, cereal grains and flowers (Francavilla & Joye, 2020). Al-Farsi et al. (2005) reported that dates fruits may be considered to be poor sources of anthocyanins compared to other fruits such as blueberries and blackberries which is rich in anthocyanin (Samad et al., 2016).

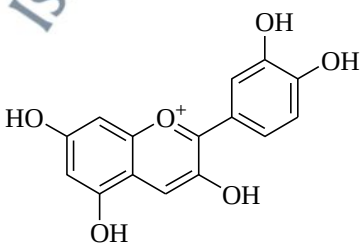
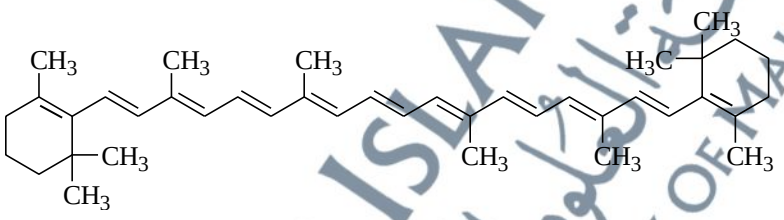


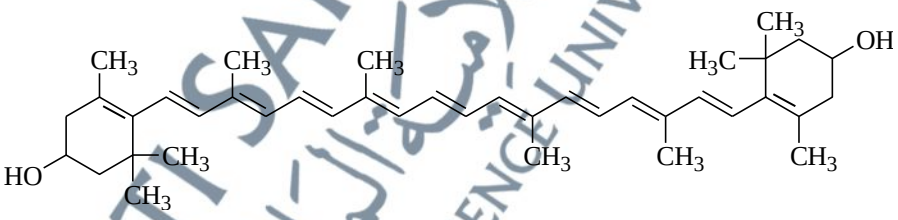
Figure 2.4: Structure of Cyanidin in Dates (Baliga et al., 2011)

2.4.3 Carotenoids

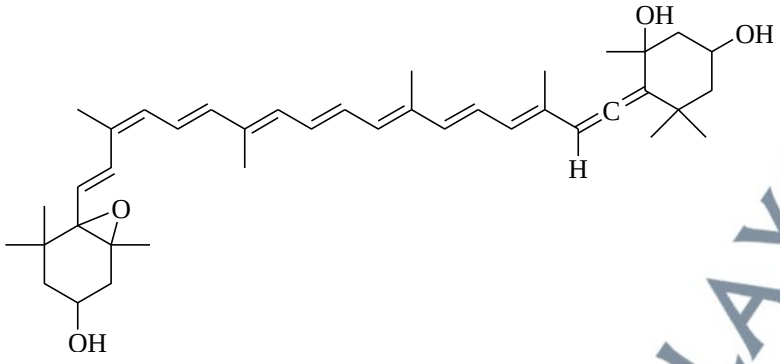
Carotenoids are a class of natural fat-soluble pigments and impart bright coloration to the plants (Baliga et al., 2011). Boudries et al. (2007) reported that dates fruits contain carotenoids; lutein, β-carotene and neoxanthin (Figure 2.5). Yellow color indicating the presence of carotenoids as the total carotenoids content in fresh dates in Khalas is the highest with 3.03mg/100g as expected because this variety has a yellow color whereas Fard and Khasab are red with total carotenoids content are 1.39 mg and 1.31 mg, respectively (Al-Farsi et al., 2005).



β-carotene



lutein



neoxanthin

Figure 2.5: Structure of Carotenoids in Dates (Baliga et al., 2011)

2.4.4 Procyanidins

Condensed tannins or procyanidins is the main precursors of blue-violet and red pigments in fruits, vegetables, nuts, seeds, flowers and barks (Baliga et al., 2011). Benmeddour et al. (2013) reported that tannins content at the final stage of fruits ripeness ranged between 600 and 2700 mg / 100 g dried weight, while at tamar stage contained 150 mg / 100g fresh weight. Moreover, condensed tannins represent the major constituents of the date's phenolic compounds and the ratio condensed tannin / total phenolic of the dates was in the range of 0.44-0.82. Figure 2.6 shows the structure of procyanidins in dates.

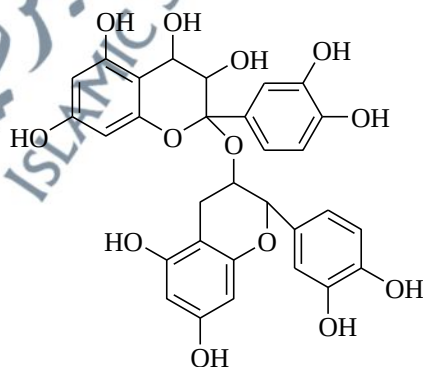
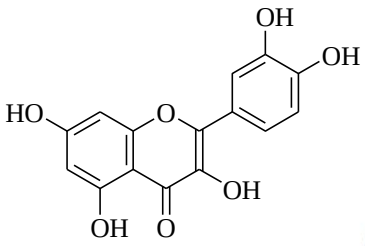


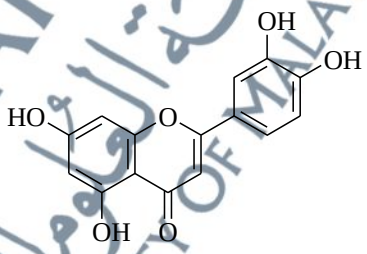
Figure 2.6: Structure of Procyanidins in Dates (Baliga et al., 2011)

2.4.5 Flavonoids

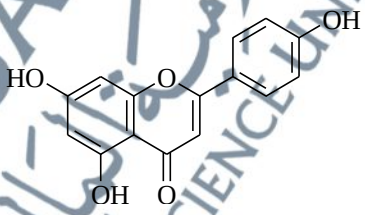
Odeh et al. (2014) stated that flavonoid is highly and significantly correlated with antioxidant capacity in dates fruit. According to Farag et al. (2014), flavonoid identified in dates fruits are quercetin, luteolin and apigenin (Figure 2.7). According to Baliga et al. (2011), some studies stated that flavonoids present in plants possess diverse health benefits, which includes antioxidant and radical scavenging activities, reduction of certain chronic diseases, prevention of some cardiovascular disorders and certain kinds of cancerous processes.



quercetin



luteolin



apigenin

Figure 2.7: Structure of Flavonoids in Dates (Baliga et al., 2011)

2.4.6 Sterols

Sterols or steroid alcohols are a subgroup of steroids with a hydroxyl group at the 3-position of the A-ring and are amphipathic lipids. Sterols in dates fruits consist

of cholesterol, campesterol, stigmasterol, β-sitosterol and isofucosterol (Baliga et al., 2011) (Figure 2.8).

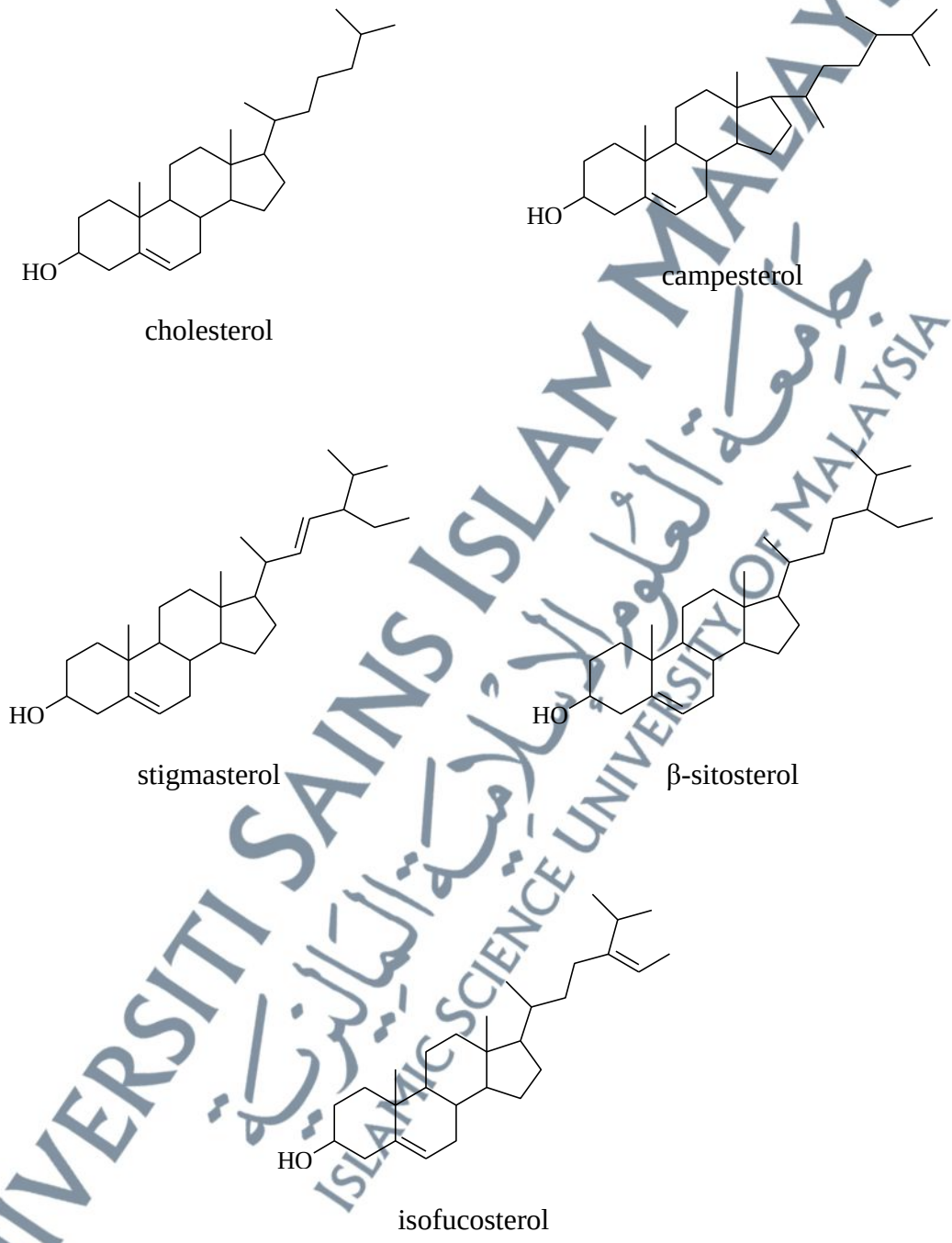


Figure 2.8: Structure of Sterols in Dates (Baliga et al., 2011)

2.5 Extraction of Chemical Compounds in *Phoenix dactylifera* L.

2.5.1 Extraction Method and Solvent Extraction

The first step to separate the targeted or untargeted chemical compounds from fruits or plant is extraction (Zhang et al., 2018b). The extracting solvent and extraction technique are significantly affected extraction yield, total polyphenol and biological activities (Hayouni et al., 2007; Trabelsi et al., 2010). According to Aris et al. (2013) which study on nutrient in Mariami dates seed oil stated that supercritical carbon dioxide (SC-CO_2) extraction method has many advantages over the conventional method such as solvent extraction due to high selectivity and diffusivity and the solvent (CO_2) is easy to separate from the extract. On the other hand, a study in determining the phytochemical composition in Tunisian dates seed and the effect of different solvents extraction was conducted by Thouri et al (2017). Another study on dates seed, Sirisena et al. (2017) used the solid phase extraction (SPE) to extract the free polyphenol. In different case, Abdallah et al. (2018) stated that QuEChERS which is stand for quick, easy, cheap, effective, rugged and safe method has many advantages over the most conventional methods for the analysis of pesticide residue. For the antioxidant extraction in dates fruits, 1,1-diphenyl-2-picrylhydrazyl (DPPH) method was used meanwhile Folin-Ciocalteu test was performed to determine the phenolic content in dates fruits (Mansouri et al., 2005; Awad et al., 2011).

Several studies used the solvent extraction such as aqueous or polar solvents to extract Tunisian dates fruits (Hammouda et al., 2013; Kchaou et al., 2016), fresh and sun-dried dates grown in Oman (Al-Farsi et al., 2005) and dates leaf and pit (Perveen et al., 2012). Therefore, there is no single extraction method may be considered universal because different target compounds requires different method of extraction in order to get higher yield (Awad et al., 2011).

The crucial part in solvent extraction is the selection of the solvent (Zhang et al., 2018b). According to Hayouni et al (2007), different group of compounds will be extracted using different solvents extraction and different techniques. In addition, Rezaie et al. (2015) stated that the efficiency of the different solvent extraction strongly depends on the matrix of plant materials and the type of extractable compounds. Hayouni et al. (2007) stated that polar solvent is the most efficient solvents to extract polyphenol in selected Tunisian fruits. Mohammedi & Atik (2011) reported that there are good correlations between phenolic and antioxidant activity. However, there is no single agreeable solvent extraction to extract chemical compounds because of the compounds with varying polarities in raw plant material (Thouri et al., 2017; Awad et al., 2011; Wijekoon et al., 2011).

Three mixture design or also known as simplex-centroid design is a type of mixtures design used to analyze the relationship involved in a process that contains several variables (Cornell, 2002). Simplex-centroid design is constructed to form a triangle with the three corners correspond to single components, the three midpoints of each side correspond to binary mixtures and the centre of the triangle correspond to ternary mixture which equal mixture of all three components (Brereton, 2003). In other studies, dates fruits were reported to have flavonoids (Farag et al., 2014), phenolics (Benmeddour et al., 2013), sterols, anthocyanins, carotenoids and procyanidins (Baliga et al., 2011) compounds which identified using liquid chromatography technique.

2.6 Metabolomics Study

Metabolomics originally refers to the sum of the pool of cell metabolites (Roux et al., 2011). Metabolome defined as quantitative measurement of low-molecular-weight of metabolite present in the biological fluid like saliva, blood and urine. In other words, it refers to the quantification of concentration of the free small metabolites pools of a biological system (Alvarez-Sanchez et al., 2010). Therefore, a broad range of endogenous and exogenous metabolites that have potential impact on the investigation of diagnosis of disease and discovery of biomarkers can be assessed (Zhang et al., 2012). Metabolomics field currently one of the fastest growing fields even though it is the most recently introduced.

2.6.1 Number of Volunteers (Sample Size)

One of the crucial parts in analysis is determining the sample size. Recently, there are no standard methods for sample size estimation in metabolomics (Nyamundanda et al., 2013). Both analyses; qualitative and quantitative require different size of the sample where qualitative analysis require smaller size of the sample than quantitative analysis (Creswell, 2012). Sample size for qualitative analysis should be enough to obtain information (Creswell, 2012) or lead to the repetition which is no additional information even after adding more sample thus, researcher should follow the concept of saturation (Fusch & Ness, 2016).

Others studies have recommended a guideline for determining the size of sample in qualitative analysis which is 30 to 50 volunteers (Morse, 1994) and 20 to 30 volunteers (Creswell, 1998). For different studies, Creswell (1998) suggested 5 to 25 volunteers and Morse (1994) recommended at least 6 volunteers. In conclusion, there are no exact guidelines to determine an appropriate sample size in qualitative research

in which, the size of the sample may best be determined by the time allotted and study objectives (Creswell, 2012).

2.6.2 Effects of Dates Fruits Consumption

Dates can be used as a nutritious food and major role in human nutrition and health (Assirey, 2015) and as excellent food ingredient (Ahmed et al., 2013). The presence of nutrients and phytochemicals including phenolics, sterols, cyanidin, carotenoids, procyanidins and flavonoids in dates fruit has been related to the health-benefits of dates fruits consumption which perfectly fit into a healthy lifestyle (El-Sohaimy and Hafez, 2010). Metabolomics investigations attempt to detect and profile the changes in metabolites, which reflect changes in metabolic pathways and may provide information concerning a disease state or the biological stress of an organism (Saude & Sykes, 2007) or for discovery of biomarker (Roux et al., 2011). According to Roux et al. (2011), metabolome consist of a set of small molecular mass organic compounds in body, which is called metabolites.

Nutritional metabolomics have great potential as an effective tool in integrating the human metabolic response to dietary interventions, nutrition and lifestyle habits. Metabolomic analysis is normally performed on biofluids, like blood, urine and feces (Savorani et al., 2013). Consumption of nutrients and bioactive compounds from food component interact with numerous targets, metabolic functions and pathways in the organism and hereby has the potential to reduce or increase the risk of diseases (Gibney et al., 2005; Salvamani et al., 2016). Targeted and untargeted metabolomics has been shown to generate global metabolic profiles (fingerprints) representative for dietary intake (Radjursoga et al., 2017). In nutrition research, the chronic effect is

relevant but in term of understanding of nutritional metabolomic data, understanding about acute dietary effect is important (Walsh et al., 2007).

According to Yasin et al. (2015), dates fruits have a potential in preventing depletion of intrinsic protection from oxidative cell damage as the fruits contain high antioxidants compounds. Moreover, Kchaou et al (2013) stated that many researchers reported that antioxidant in dates fruits is effective against gastric ulcer (Al-Qarawi et al., 2005), hepatotoxicity (Saafi et al., 2011) and could reduce the risk of major chronic health problems including cancer, atherosclerosis, cardiovascular disease (Al-Yahya et al., 2016), neurological disease and coronary heart disease and diabetes (El-Hadrami and Al-Khayri, 2012; Zhang et al., 2013).

Other studies found that consumption of dates fruits could reduce the need for induction and augmentation of labour (Al-Kuran et al., 2011), help to control bowel movement and constipation due to consist of high fiber that lead to easy digest, containing nutrient such as potassium to help in balancing human nervous system, phosphorus and calcium for bone strength and growth and natural sugars to give instant energy to human body (El-Sohaimy and Hafez, 2010; Aslam et al., 2013).

2.6.3 Urine Sampling and Storage Effect

Urine is a liquid produced by the kidneys that contained unwanted compounds in the body which is the waste from the end products of cellular processes (Denkert et al., 2012). Urine collection is persistent, does not need patient preparation and considerably improves compliance. Over than 10 years, urine was used as the specimen in metabolomic study as it is easily obtained, readily available, simple and lower protein content so it just need simple sample pre-treatment and a rich source of biomarkers (Zhang et al., 2012). Diet gives notable change in urine composition

because the presences of exogeneous metabolites like phytohemicals coming from food component affect the human metabolism (Gibney et al., 2005). Furthermore, according to Kind et al. (2007), the composition and quantity of metabolites in urine is in response to metabolic state, medication and diet.

Sample collection for urine metabolomics is a sensitive, convenient and non-invasive technique, readily available, easy to obtain in large volume and can be performed either as spot urine (random), 24 hours or as first morning void (Witte et al., 2009; Savorani et al., 2013; Rodriguez-carrasco et al., 2014). This sampling mode is mandatory when the excretion of urine is monitoring after intake of food or drug. In this case control urine, that is urine before intake of food or drug is required to compare the changes of urine after intake. Example of timed samples studied by Villalba (2010) based on targeted analysis of phenols from olive oil in urine. In the study, samples were collected during the wash out period (first void spot urine in the morning) and after oil ingestion at three different periods (0-2 hours, 2-4 hours and 4-6 hours).

However, the 24 hours sampling collection was aimed to eliminate large variability in metabolite profile. The common example for 24 hours is sampling urine day and night. The 24 hours urine collection allowed the researchers to investigate the activity of gut microbiota in healthy people during day and night as it indicates the presence of biomarkers. The result shows that there is obvious change in the activity of gut microbiota with the circadian rhythm (Slupsky et al., 2007; Peralbo and Castro, 2012). Although 24 hours sampling can be quite burdensome, the excretion picture that they provide is usually complete.

The effect of sample storage has been widely study in biofluids. The sampling strategy and storage of biofluids is really important because of metabolic activity

during sampling and storage must be stopped, therefore the temperature must be reduced during sample preparation (4°C) and storage (-80 °C) (Alvarez-Sanchez et al., 2010). Lauridsen (2007) found that the storage of human urine samples at temperature -25 ºC give no change for 26 weeks for ^{1}H-NMR fingerprints. When samples stored at 4 °C without preservative added, there is acetate formation due to microbial contamination.

2.6.4 ^{1}H-NMR-Based and MS-Based in Metabolomic Analysis

There are two metabolomics based analysis; nuclear magnetic resonance (NMR-based) and mass spectrometry (MS-based). Among these analytical platforms, ^{1}H-NMR-based metabolomics is widely used compared to MS-based metabolomics since it can simultaneously display resonance peaks resulting from hundreds of metabolites (Gu et al., 2007; Duarte et al., 2014) and requires little or no sample preparation (Lehtonen et al., 2013; Gu et al., 2007), thus making it suitable for high-throughput analysis (Nguyen et al., 2016). Furthermore, Nguyen et al (2016) stated that the advantages of ^{1}H-NMR-based metabolomics approach are highly reproducible and non-destructive.

The combination of ^{1}H-NMR-based metabolomics with multivariate statistical analysis is a powerful approach for the study of metabolomics (Gu et al., 2007), which may help to uncover features of the biomarkers that are still hidden (Fanos et al., 2014) and allows the explanation of changes in metabolic pathways under different psychological conditions (Brennan, 2014). Meanwhile, many researches using MS-based analysis due to the high sensitivity and selectivity. For instance, liquid chromatography mass spectrometry (LCMS) could be benefited from lower detection limits and improved MS data quality due to reduced background noise (Naz et al.,

2014) and Niu et al (2014) stated that gas chromatography mass spectrometry (GC-MS) gives high resolution chromatographic separation and wide applicability through derivatization.

Since 1973, Gas Chromatography Mass Spectrometry (GC-MS) and in about 5 years later, Liquid Chromatography Mass Spectrometry (LCMS) has been used in clinical biochemistry especially for urine analysis (Kind et al., 2007). Recently, this technology has been developed for metabolomics study in animals as well as human with the goal of toxicology and biomarker discovery, disease diagnosis and nutrition study. GC-MS is preferable for the analysis of non-polar analytes like synthetic organic compounds and natural products (Garcia and Barbas, 2011). Mass spectrometry- based has receiving great attention among researchers over the world as it combines the sensitivity of MS analysis with the possibility to identify metabolites in the usually complex sample that are analyzed. Gas chromatography mass spectrometry (GC-MS) is very useful in identifying metabolic changes as shown in large number of publications (Josson et al., 2005). In order for a compound to be analysed by GC/MS it must be sufficiently volatile and thermally stable (Chromacademy, 2012).

According to Farag et al. (2014), abundant researches have been done to discover the possible technique to detect hundreds of metabolites of living organisms. To compare with conventional liquid chromatography (LC), Ultra Performance Liquid Chromatography (UHPLC) gives greater peak separation using standard setting methods for LC. Particularly, LC coupled with quadrupole time of flight-mass spectrometry (QToF-MS) with electrospray ionization (ESI) is a well-known

technique used to perform untargeted metabolomics in plant extracts (Vaclavik et al., 2011) with excellent sensitivity (Kronstrand et al., 2014).

Additionally, Jandric et al. (2014) who studied the adulteration in juice stated that there are researches on various analytical methods which are from simple to more complex techniques such as capillary electrophoresis (Saavedra et al., 2001), spectroscopy (Cuny et al., 2008) and liquid chromatography or gas chromatography coupled to several type of detectors (Ehling & Cole, 2011). In summary, although there are many analytical techniques used in metabolomics analysis, combination of NMR-based and MS-based analysis is an ideal instrument to be used in metabolomic research.

2.7 Principal Components Analysis (PCA)

2.7.1 Application of PCA in Metabolomics Study

PCA is a non-supervised multivariate data analysis technique which commonly first applied in non-targeted metabolomic study (Naz et al., 2014; Yao et al., 2014). According to Savorani et al. (2013), multivariate dataset of metabolomics study usually analysed using PCA as it may analysed the large scale dataset simultaneously. This is because, PCA can extract more detailed about the information of metabolic beyond the visual inspection (Gu et al., 2007) and more time and cost effective when spectroscopy data coupled with chemometric (Uarrota et al., 2014). Generally, multivariate data in metabolomics study is a large data that difficult to analyze manually thus, PCA was used to visualize the multivariate data in grouping trends (Ramakrishnan et al., 2016; Feng et al., 2016; Yao et al., 2014) and to reduce the complexity of the data sets (Park et al., 2016). Therefore, researcher could

comprehend their multivariate data as the PCA is highlighting the similarities and the differences within the samples (Naz et al., 2014).

According to Shen et al (2019), the PCA was constructed to visualize the comparison of metabolic response of the fish fed with four different percentage of taurine supplementation. Other metabolomics study using PCA to differentiate between two species of *Artemisia* (Liu et al., 2010), to interpret the metabolites changes in prostate cancer (Struck-Lewicka et al., 2015), to distinguish between two groups of urine metabolites of anti-stress effect (Feng et al., 2016) and to investigate the pattern of dietary intake in metabolomics study (Gibbons et al., 2017).

In fruits metabolomic study, Aguiar et al (2014) who is study on the jelly palm fruits and Lee et al (2010) who is study on guava fruits were used the PCA to investigate the chemical composition in the fruits at different stages of maturity. Meanwhile in different study, PCA has been applied by Wang et al (2003) to distinguish the different NMR spectral editing techniques in liver study. Niu et al (2014) stated that data matrix of eight different programs commonly used in metabolomics study were subjected to analysis by PCA to highlight compounds that differed between groups of samples.

2.7.2 The Theory of Principal Components Analysis (PCA)

PCA is the general name for a technique which uses sophisticated underlying mathematical principles to transform a number of possibly correlated variables into a smaller number of variables called principal components. PCA is often used to analyze large data sets. It is the most widespread multivariate chemometrics technique. In general terms, PCA uses a vector space transform to reduce the dimensionality of large data sets. Using mathematical projection, the original data set,

which may have involved many variables, can often be interpreted in just a few variables. Then, trends, patterns and outliers can be spotted in the data (Richardson, 2009). Figure 2.9 illustrates the example of information arrangement in data matrix.

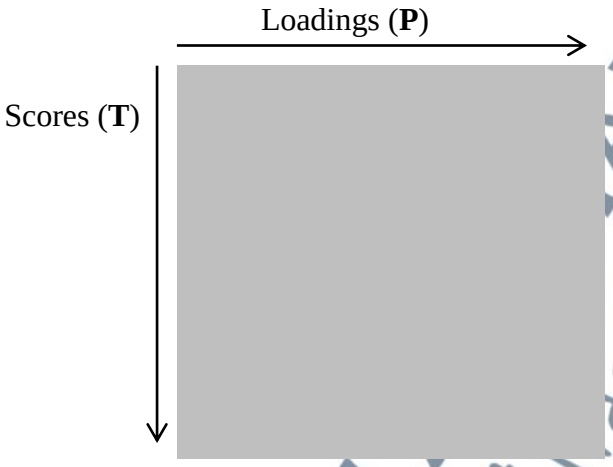


Figure 2.9: Example of Information Arrangement in Data Matrix

The samples from multivariate data are arranged into the data matrix (scores) meanwhile the parameters are on the loadings section. Tambellini et al. (2013) stated that unsupervised principal component analysis (PCA) is a tool which provides a multivariate overview of the data based on the underlying variance between the metabolite profiles of the samples without specifying the different sample types meanwhile Yang et al. (2017) stated that PCA used to convert the multivariate data into lower dimensions, which is providing a compact data in terms of plot visualization.

The formula for PCA is as follow:

$$\boldsymbol{X} = \boldsymbol{T.P} \qquad (2.1)$$

Where

T is called the scores, and has many rows as the original data matrix;

P is the loadings, and has as many columns as the original data matrix;

The number of columns in the matrix **T** equals the number of rows in the matrix **P** (Figure 2.9). From the statistical data, it is easy to convert from numbers to graphs. There are many ways of visualizing PCs. The first two PCs that are widely used are scores plots and loadings plots (Brereton, 2003). First, a new set of orthogonal coordinate axes will be identified from the data. It is done by obtaining the (least-square) line of best fit through the plotted data. The axis is called the first principal component of the data. Then, by using orthogonal projection, the coordinates will be mapped to the axis. After that, a second principal coordinate (axis) which is both orthogonal to the first PC will be obtained. Now, the coordinates can be mapped into the two orthogonal principal components. This type of diagram is called scores plot (Richardson, 2009).

The loadings plots are the relation between the original variables and subspace dimensions. If the scores plots are used for interpreting relation among samples, then loadings plots are used for interpreting relations among the variables. Meanwhile, Figure 2.10 and Figure 2.11 show the example of scores plot and loadings plot respectively (Brereton, 2003).

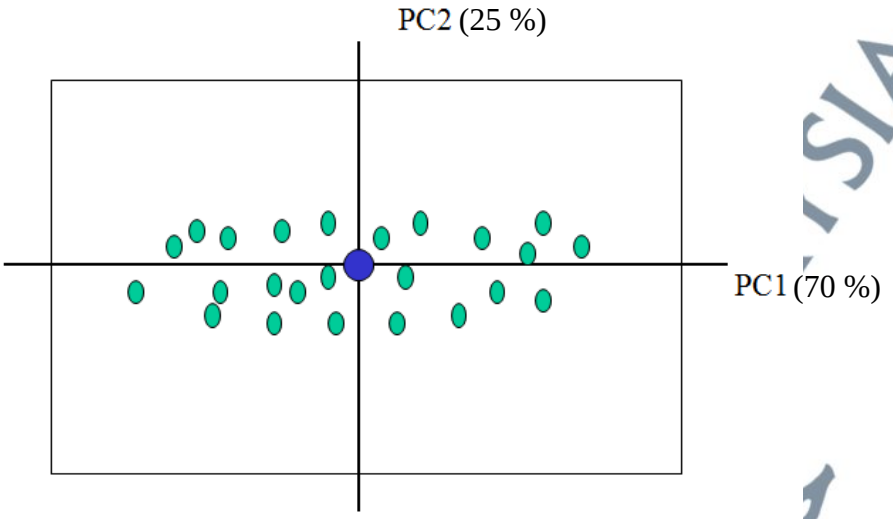


Figure 2.10: Scores Plot (Brereton, 2003)

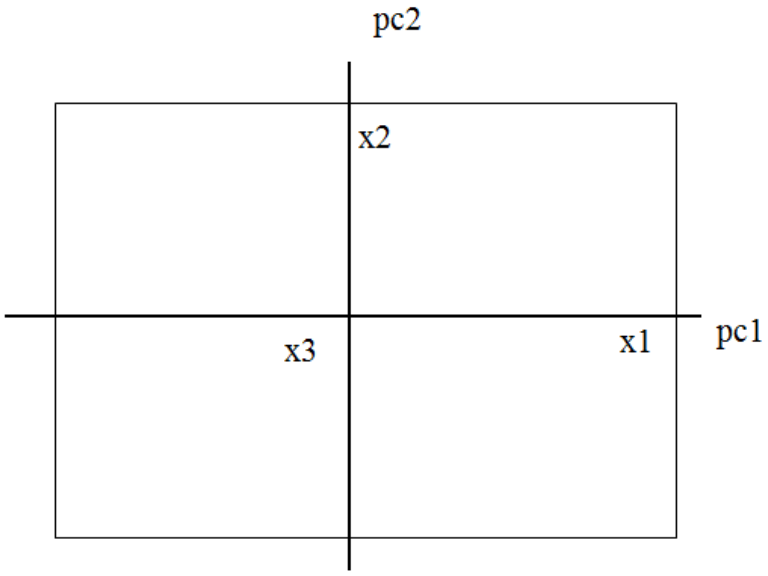


Figure 2.11: Loadings Plot (Brereton, 2003)

2.7.3 Pre-Processing Methods in PCA

There is no clean data of metabolites that provided by analytical instrument. Pre-processing of raw data must be done to produce a practical data matrix in variety of techniques (Yi et al., 2016). Scaling or data preprocessing can give significant

influence on the result. In metabolomics, it is mainly the method for preprocessing that influences the outcome. In addition, another common technique for data scaling is standardization. In some situation, standardization (or similar types of data preprocessing) is essential. For example in metabolomics, where the sample commonly used is urine, in which have abundant of metabolites but their variation is not significant among all samples. If the concentration of the minor compounds were changed, it might have a significant relationship to the underlying biology. Otherwise, PCA will be dominated by the most intense metabolites if standardization is not implemented.

For example, Yao et al (2014) used the mean centering for data pretreatment before performing PCA while other study was used the self-scaled on the data matrix (Aguiar et al., 2014), mean-centered scaling (Park et al., 2016; Lee et al., 2010; Wang et al., 2003) and Pareto-scaling (Ramakrishnan et al., 2016). However, Niu et al (2014) stated that the data matrix was subjected to several preprocessing steps without mentioning the details.