Biochemical and genetic analyses of *Jatropha curcas* L. seed composition

Wei He

PhD Thesis

University of York

Department of Biology

August 2011

Abstract

Diminishing worldwide fossil fuel reserves coupled with the negative impact of their use on the environment has led to increased research and development of renewable energy sources. Renewable liquid biofuels are in demand for the transport sector, particularly if they can be used directly in existing internal combustion engines. *Jatropha curcas* L. is a perennial plant which belongs to the *Euphorbiaceae* family. *J. curcas* seeds contain about 30% oil which is suitable for biodiesel production, and therefore it has received global interest as a source of biofuel. However, to date *J. curcas* has not been put through any stringent breeding program for traits improvement and thus has not reached its full potential. Improving seed yield, seed oil quality and quantity is necessary for large scale biodiesel production, and developing other by products will add to the economic value of this crop. Apart from oil, *J. curcas* seeds also contain a high percentage of proteins, which makes the seed meal potentially useful as animal feed. However, seeds from most of the current global *J. curcas* plantations in Asia and Africa are characterized as non-edible, due to the existence of a few toxins or antinutrients. Phorbol esters have been considered as the main toxic agent. Edible provenances exist in Mexico which are devoid of phorbol esters. Among the other toxins, curcin (a type I ribosome inactivating protein) levels have not been reported for the edible and non-edible varieties.

To improve our understanding of *J. curcas* natural variation and biochemical composition, seed samples were collected from a variety of locations in Madagascar, Mexico, and purchased from five other countries. Seed oil content, fatty acid composition and phorbol esters content were measured to establish the diversity in these traits. Seed oil content and fatty acid composition was found to vary in seeds collected from different sites, and oleate and linoleate composition were found to correlate strongly with cultivation site temperature indicating the importance of environmental conditions for the production of an optimal feedstock for biodiesel. Phorbol esters were found to be present in all seed samples originating from outside Mexico, and in the Mexican provenance Rosario Chiapas. All other Mexican samples lacked phorbol esters. This suggests that the presence of phorbol esters is a qualitative trait. AFLP analysis revealed that most genetic variation was present in

Mexican samples, with all material originating from outside Mexico showing very limited genetic diversity. Edible samples and non-edible samples were found to be genetically distinct, with the edible samples forming a single cluster. The large amount of variation in oil quantity and quality in Madagascan samples, together with the limited genetic diversity in these samples, implies that *J. curcas* seed oil is largely influenced by environmental factors.

Seed curcin levels were determined in the edible and non-edible varieties. The results showed that curcin levels are equally abundant in both varieties. This demonstrates that curcin is not playing any significant role in determining seed edibility and is consistent with the predominant role understood to be played by phorbol esters in determining this trait. The spatial and temporal expression of different curcin genes was further examined. Four curcin genes showed different patterns of expression with seed and leaf specific patterns of expression being identified. Further analyses revealed that *CURCIN2* is induced in mature leaves in response to various abiotic stresses. Furthermore it appears that the induction of *CURCIN2* in response to wounding is regulated via the JA (jasmonic acid) signalling pathway.

Together these results represent a valuable addition to the knowledge base underpinning the development of *J. curcas* as an industrial crop through molecular breeding.

Table of contents

List of figures

List of tables

Publications

HE, W., KING, A. J., KHAN, A. W., CUEVAS, J. A, RAMIARAMANANA, D. & GRAHAM, I. A. 2011. Analysis of seed phorbol-ester and curcin content together with genetic diversity in multiple provenances of *Jatropha curcas* (L.) from Madagascar and Mexico. *Plant Physiology and Biochemistry*, 49, 1183-1190.

KING, A. J., HE, W., CUEVAS, J. A., FREUDENBERGER, M., RAMIARAMANANA, D. & GRAHAM, I. A. 2009. Potential of *Jatropha curcas* as a source of renewable oil and animal feed. *Journal of Experimental Botany*, 60, 2897-2905.

Acknowledgements

I am truly indebted and thankful to Professor Ian Graham for his supervision, enthusiasm, inspiration and encouragement for my project. I have enjoyed tremendously working in his lab during these past four years. I have learned a great deal from his positive thinking, which will vastly benefit me in my following endeavours. I would like to show my gratitude to Dr Andrew King for his co-supervision, knowledgeable and constructive advice, tolerance and patience in answering my many questions. I appreciate all the time Professor Ian Graham and Dr Andrew King spent helping me to become an independent researcher.

I would like to thank Dr Fabian Vaistij very much for his co-supervising, sacrificing his research time to provide me with many useful discussions, and bearing with my disorder in the lab. Thanks to Professor Dianna Bowles and Dr Michael Schultze for their help at training committee meetings. My thanks to the many great people in the Graham lab and also in CNAP who have helped me in my studies, including Mrs Judith Mitchell, Dr Tony Larson, Dr Vasilios Andriotis, Dr Anuja Dave, Dr Zhesi He, Dr Yi Li, Dr Awais Khan and Ms Valeria Gazda.

I am grateful to the ORSA and CSC scholarships for funding my project.

I would like to thank my parents for their constant and unconditional love and support. Sincere and earnest thanks to my wife Ya Wen for her dedication to accompanying me in York for my studies, and her infinite love, care and encouragement. To my dear friends, including a few of the names mentioned above, who made my stay in York enjoyable and worthwhile. This dissertation would not have been possible without them.

Finally, this thesis is also to my dearly beloved grandfather, who passed away in the year I started my Ph. D. Whenever I felt weak during these past four years, his spirit was always there to encourage me. To my dog, who left us before just I finished this thesis. He had been a family member for ten years and gave us precious and immense unforgettable memories. They will always be remembered.

Author’s declaration

I declare that I am the sole author of the work in this thesis and that it is original except where indicated by special reference in the text. No part of this degree has been submitted for any other degree to any other institution.

Chapter 1: Introduction

1.1 The status quo of worldwide fossil fuels

Fossil fuels, which consist of oil, gas and coal, have provided the energy needs for population growth and have been the driving force behind industrial development for more than a century. Since the oil crisis in 1973, the global oil price has increased dramatically (from USD 2.8 per barrel in 1973 to USD 93.0 per barrel in 2011; http://www.oil-price.net/, accessed on 23rd June 2011). However this has not reduced the worldwide demand for oil. Likewise, this is also true for gas and coal consumption (Figure 1.1). The trend of this increasing demand for fossil fuels will remain and it is predicted that by 2030, the energy consumption from fossil fuels will grow by approximately 30% (British Petroleum, 2011).

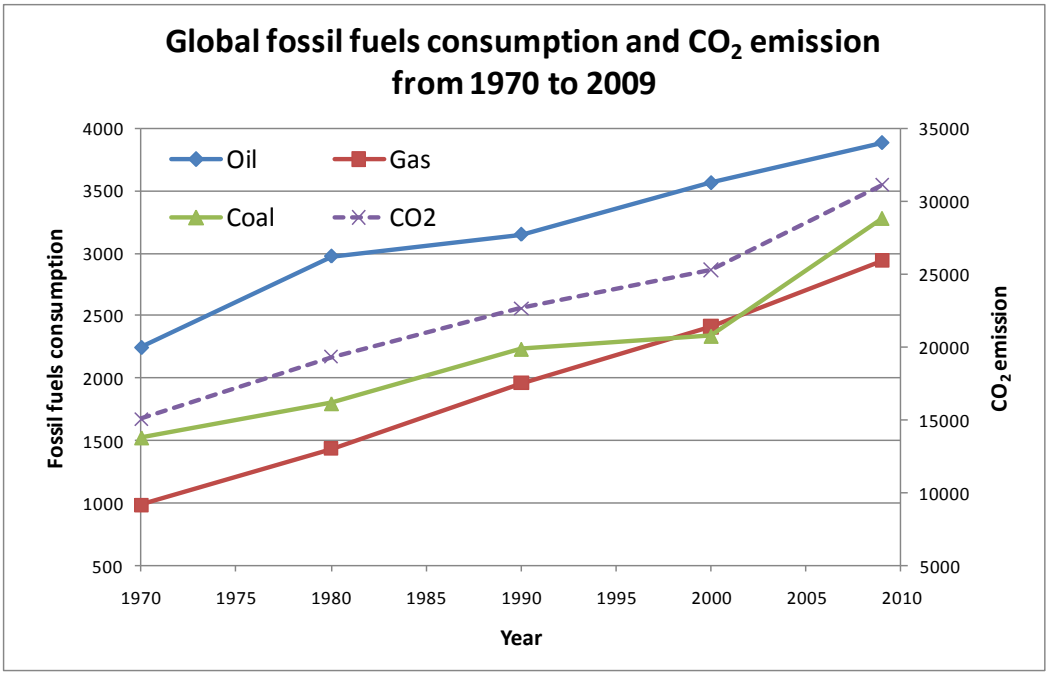


Figure 1.1 Global fossil fuel consumption and CO_2 emission from 1970 to 2009. Data represent the annual consumption of oil (million tonnes), gas (billion cubic metres), coal (million tonnes oil equivalent) and the annual emission of CO_2 (million tonnes). All data collected from "BP Statistical Review of World Energy 2010" (British Petroleum, 2010).

At the end of 2009, the global fossil fuel reserves stood at 181.7 billion tonnes of oil, 187.5 trillion m^3 of gas, and 826 billion tonnes of coal (Table 1.1). Given the current reserve-to-production ratio, fossil fuel reserves will only last another 46 years for oil, 63 years for gas and 119 years for coal (Table 1.1). As demand for fossil fuels is

increasing—particularly in developing countries such as China and India, it is likely that these fossil fuel reserves will be depleted sooner (British Petroleum, 2010). The uneven distribution of energy resources among countries, the limited supplies and the consequent fluctuating fuel prices have led to significant vulnerabilities in energy security in countries which rely heavily on fossil fuels imports.

Table 1.1 Calculated reserves of fossil fuels at the end of 2009 and the reserve-to-production (R/P) ratio. Data collected from "BP statistical review of world energy 2010" (British Petroleum, 2010).

Type of fossil fuels	Proved reserves at the end of 2009		R/P (reserves-to-production) ratio
Oil	181.7	billion tonnes	46
Gas	187.5	trillion (10^{12}) m^3	63
Coal	826.0	billion tonnes	119

The use of fossil fuels is also problematic due to the fact that combustion of these fuels releases the greenhouse gas CO_2. From 1970 to 2009, the emission of CO_2 almost doubled (Figure 1.1). The combustion of fossil fuels is thought to be the main contributor to the observed steady increase in the Earth's atmospheric CO_2 concentration from 315 ppm (parts-per-million) in 1958 to 390 ppm in 2010 (Mauna Loa observatory data, http://www.esrl.noaa.gov/gmd/ccgg/trends/). The increased global CO_2 concentration has been suggested as one of the main causes of global warming—according to the Intergovernmental Panel on Climate Change, the global surface temperature increased by 0.74 ± 0.18 °C (1.33 ± 0.32 °F) from 1905 to 2005 (Solomon et al., 2007). Combustion of fossil fuels also releases a number of pollutants, including particles, NO_x and SO_x, which react with water molecules in the atmosphere and cause acid rain (Likens and Bormann, 1974). NO_x is also considered to cause photochemical smog (Seinfeld et al., 1998).

1.2 Developing renewable energy sources

Fossil fuels accounted for 86.4% of the total energy production, which reflects our strong dependence on them (Figure 1.2). Nuclear energy contributes 5.6% to the total energy production. Although this source of energy releases less CO_2 than fossil fuels (International Atomic Energy Agency, 2009), and provides an opportunity for countries poor in fossil fuel reserves to achieve greater energy security, the use of

nuclear energy still causes problems. Disposal of nuclear waste containing radioactive material has potential risks to the environment; a number of nuclear accidents, including the Chernobyl incident in 1986 and the recent nuclear crisis in Japan, highlighted the health and safety issue.

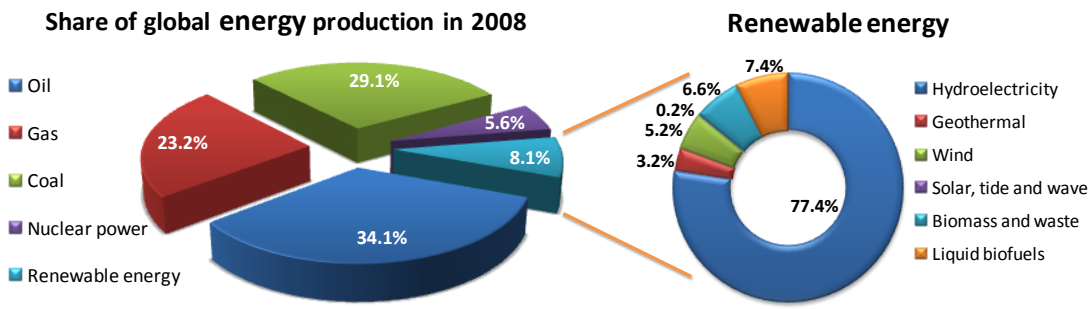


Figure 1.2 Share of global energy production in 2008. Data from International Energy Statistics, U.S Energy Information Administration (http://www.eia.doe.gov/).

Due to the finite supply of fossil fuels, and the fact that their usage raises a number of environmental, political and socioeconomic problems, research into alternative renewable sources of fuel is currently very active. In 2008, renewable energy accounted for 8.1% of the total energy production (Figure 1.2). The main contribution is from hydroelectricity (77.4% of renewables or 6.3% of the total energy production), with the remaining renewable energy sources, which include geothermal, wind power, solar power, tide and wave power, biomass and waste, and liquid biofuels, accounting for just 1.8% of the total energy production. However, there has been a steady increase in the contribution of renewables (excluding hydroelectricity) since the start of the 21st century, and it has been predicted that by 2030, renewables will account for over 5% total energy production (Figure 1.3A); this accounts for 18% of the growth in total energy by 2030 (Figure 1.3B).

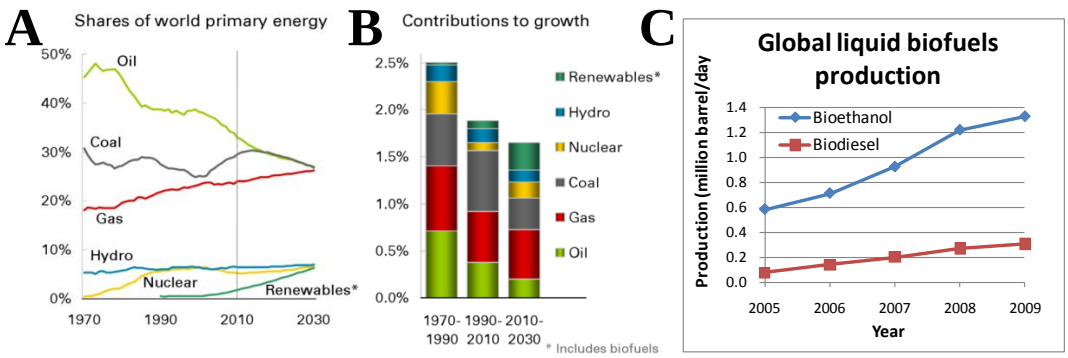


Figure 1.3 Historic and projected share of world primary energy (A), contributions of different sectors to growth (B), and global liquid biofuels production: bioethanol and biodiesel (C). Figure A and B are from BP Energy Outlook 2030 (British Petroleum, 2011). In Figure B contributions to growth are shown on yearly basis. In Figure C data were collected from International Energy Statistics, U.S Energy Information Administration (http://www.eia.doe.gov/). Production is shown in million barrel of oil equivalent/day.

1.3 Renewable energy from biomass and the production of liquid biofuels

For thousands of years, biological material, including plants, animals and microbes, has been providing food, energy and other products (e.g., rubber, paper, medicines) for the needs of human beings. The term biomass refers to any living or recently living organic matter from nature, such as wood, agricultural crops and their by-products, food processing waste and municipal refuse. Traditionally, man has used energy directly from biomass by incineration, which is an efficient process. However, there is significant demand from the automotive industry for liquid fuels. Liquid biofuels can be produced from biomass by the fermentation of sugars into alcohols (bioethanol and biobutanol), or the transesterification of triacylglycerol to produce mono-alkyl esters (biodiesel). Alternatively, a number of thermochemical processes have been developed for conversion of plant biomass into fuels, such as liquefaction, pyrolysis and gasification (Pandey, 2009).

Liquid biofuels are made from feedstocks which are from nature by direct or indirect assimilation of carbon, and therefore the use of liquid biofuels contributes less to the net CO_2 emission than fossil fuels. In addition the feedstocks are renewable and relatively more equally distributed geographically. Thus substituting fossil fuels with liquid biofuels in part can increase energy security and benefit the economy. Bioalcohol or biodiesel blended with petroleum helps with its complete combustion

and therefore results in improved engine performance. This reduces the tailpipe emissions including particles, sulphur and nitrogen oxides.

As a newly developed energy source, liquid biofuels only take a small portion of the world total energy production (7.4% of renewables or 0.6% of the total energy production, Figure 1.2), predominantly by bioethanol (Figure 1.3C). Nevertheless, based on the above strengths, it is foreseeable that the global production of liquid biofuels and the related research interests will keep growing and underpin economic growth in the next few decades. Global production of biodiesel grew by 4-fold from 2005 to 2009, and it is 2.3-fold for bioethanol (Figure 1.3C). It is predicted that the biofuels production will exceed 6.5 million barrels per day (Mb/d) by 2030, up from 1.8 Mb/d in 2010 (British Petroleum, 2011).

1.3.1 Bioethanol and biobutanol

Bioethanol and biobutanol can be used as alternatives to petrol (gasoline). The sugars required for the fermentation process is most readily obtained from crops rich in either sugar or starch. Brazil and the USA are the biggest producers of bioethanol using sugarcane and corn, respectively (Brittaine and Lutaladio, 2010). Bioalcohols produced from these crops are commonly referred to as first generation biofuels. Second generation biofuels are being developed, including bioalcohols produced from lignocellulosic biomass. Typical lignocellulosic biomass includes dedicated biomass crops (e.g. poplar), trees and grasses, agricultural waste (e.g. sugarcane bagasse and corn stover) and industrial woody waste, which can be converted into alcohols, but requires pre-treatment to break lignocellulosic material into fermentable sugars (Pandey, 2009). Physical and chemical processing are commonly used in the pre-treatment of lignocellulosic biomass, but more efficient biological pre-treatment approaches are becoming developed (Pandey, 2009). Though currently more complex and costly, bioalcohols production from lignocellulosic biomass has many advantages over first-generation bioalcohols. Production from agricultural wastes does not require additional land, and the use of dedicated biomass crops results in less competition with food production.

Bioethanol is the most common bioalcohol, typically used as a low percentage blend such as E10 to E25 (i.e, containing 10% or 25% ethanol) (Pandey, 2009). With the development of "flex-fuel" vehicles, higher percentage blend such as E85 can be used. In Brazil, this has been a real success; from 2000 to 2009, more than 50% of the light vehicles sold were fuel-flex (de Freitas and Kaneko, 2011). Biobutanol is not commercially available at present, but it has significant advantages over bioethanol. It has higher energy efficiency, easier blending property with gasoline (since it is less polar than ethanol) and requires no specially adapted gasoline vehicles [U.S Department of Energy, the Alternative Fuels and Advanced Vehicles Data Center (http://www.afdc.energy.gov/afdc/fuels/emerging_biobutanol.html)]. Biobutanol is being developed by a few industrial groups, such as DuPont and BP.

1.3.2 Biodiesel

Biodiesel, as the name implies, can be used as an alternative to petrodiesel. It is made from the transesterification of triacylglycerides obtained from vegetable oils or animal fats with short-chain alcohols (Srivastava and Prasad, 2000). This transesterification can either be achieved chemically or enzymatically using lipases. This process considerably reduces the viscosity of the oil to a level similar to that of petrodiesel (Srivastava and Prasad, 2000). Many plant oils, such as soybean, sunflower and rapeseed are currently used as feedstocks for biodiesel production. Currently the USA is the largest producer of biodiesel using soybean oil as the feedstock, while in Europe rapeseed oil is widely used (Brittaine and Lutaladio, 2010). However, similar to their role in bioethanol production, food crops are not the most desirable feedstock for biodiesel as they can compete directly with the food supply. Thus many studies have been carried out to look for plants bearing non-edible oil as potential biodiesel feedstocks.

1.4 *Jatropha curcas* L. and its potential as a feedstock for biodiesel production

1.4.1 Context and general description of J. curcas

Jatropha curcas L. (commonly named physic nut or purging nut) is a perennial plant of the spurge family (*Euphorbiaceae*). The plant, which is a shrub or small tree, can

grow up to 5-8 metres (Heller, 1996). Generally, *J. curcas* grows in tropical and sub-tropical temperatures, with annual precipitation of 200 mm-1500 mm (Brittaine and Lutaladio, 2010). A typical *J. curcas* plant is shown in Figure 1.4 and the botanical description of *J. curcas* is summarized in Table 1.2. *J. curcas* is propagated both generatively (from seeds) and vegetatively (from cuttings) (Brittaine and Lutaladio, 2010, Heller, 1996). Free-draining sands and loams with no water logging are preferred for the growth of *J. curcas* (Heller, 1996).

J. curcas is thought to originate from Mexico or Central America, and was introduced to Africa and Asia by the Portuguese sailors in the 16th century (Heller, 1996, King et al., 2009). Since then it has spread widely spanning the whole tropics (King et al., 2009) with more than 85% of the *J. curcas* plantation in Asia, mainly India, China, Myanmar and Indonesia (Brittaine and Lutaladio, 2010). The exact worldwide plantation of *J. curcas* is unclear. Based on 242 identified projects and 176 expert interviews worldwide, GEXSI (The Global Exchange for Social Investment) estimated that *J. curcas* plantation would reach 5 million ha in 2010 and increase to 12.8 million ha by 2015 (Renner et al., 2008).

Historically *J. curcas* was used as a hedge plant to keep out grazing livestock and to provide oil for lighting lamps and making soap. *J .curcas* oil was also used as insecticide or molluscicide (Brittaine and Lutaladio, 2010). Medical uses of *J. curcas* have also been reported but are mostly restricted to indigenous areas (Heller, 1996).

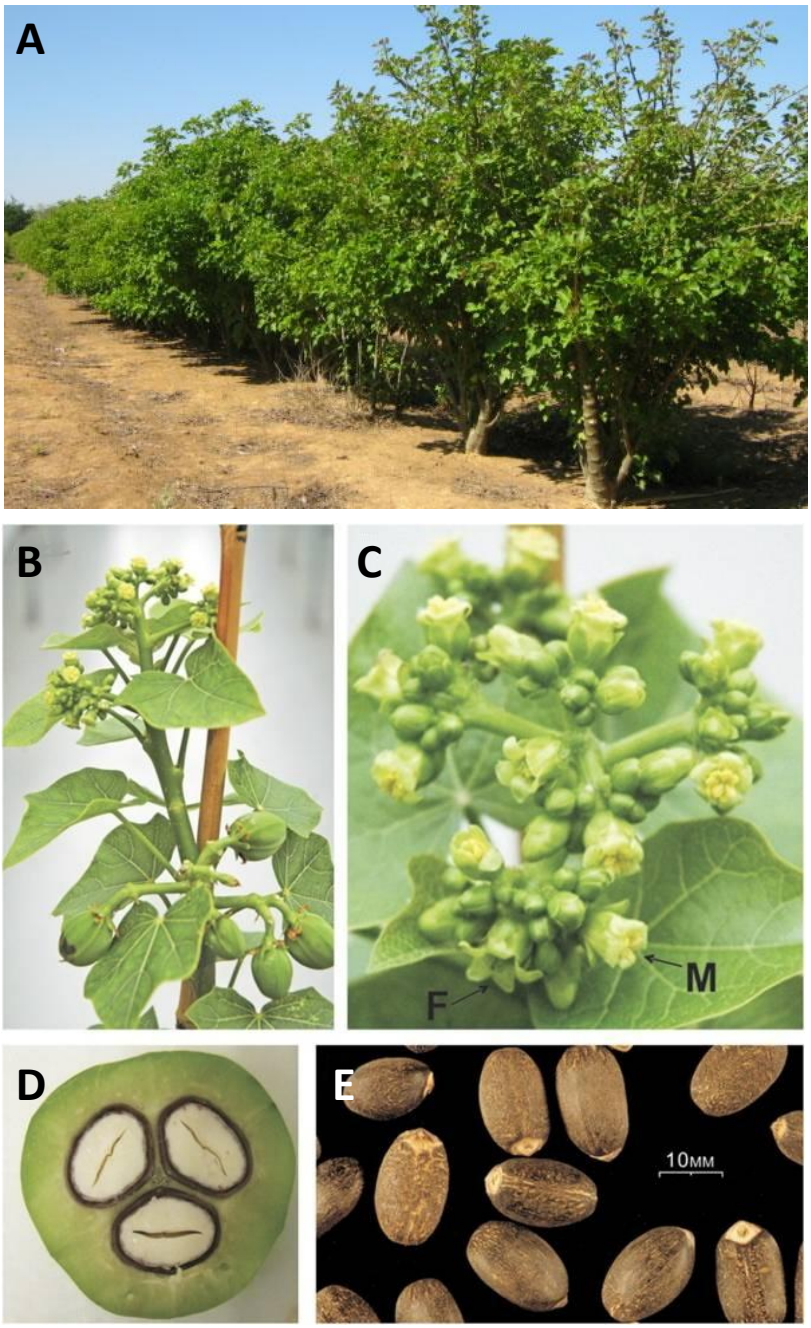


Figure 1.4 Images of *J. curcas*. A: *J. curcas* plants. B: *J. curcas* plant with both flowers and developing seed pods. C: *J. curcas* inflorescence containing both male staminate flowers (M) and female pistillate flowers (F). D: Cross-section of a *J. curcas* seed pod containing three developing seeds. E: Mature seeds of *J. curcas*. Picture A is from http://www.biofuelsrevolution.com/images/grass-stains/Jatropha-plantation-in-Brazil.jpg. Pictures B-E are from Dr. Andrew King (King et al., 2009).

Table 1.2 Botanical description of *J. curcas*. Summarized from Heller (1996), Divakara et al (2010) and Brittaine and Lutaladio (2010).

Roots	One tap root and four lateral roots, with many secondary roots; no tap root is formed from vegetative propagation.
Leaves	Leaves are green to pale green arranged alternatively; leaves are 5-7 shallow lobed, with a length and width of 6 to 15 cm; contains white latex.
Flowers	The plant is monoecious and flowers are unisexual; the inflorescence is formed on the terminal of branches, in auxiliary paniculate polychasial cymes; normal male/female flowers ratio is 10-20:1; flowers are in greenish yellow colour.
Pollination	Pollinated by insects; the plant is fully self-compatible.
Fruits	Fruits are trilocular; exocarp is green but becomes dry and turns into brown after ripening; each fruit has normally three seeds.
Seeds	Seeds are ellipsoidal and in black colour, 2 cm long and 1 cm thick; the caruncle is small; dormancy is induced by changes in rainfall and temperature/light.
Growth	The plant shows articulated growth; has straight stalk and thick branchlets; the stems and branches are hollow soft wood, containing white latex; has a life expectancy up to 50 years.

1.4.2 J. curcas as a choice of biodiesel feedstock

As a perennial plant, *J. curcas* has a deeper root system than annual crops, which helps to store more carbon, maintain soil quality and manage water and nutrients more conservatively (Cox et al., 2006, Glover et al., 2007). Much of the interest in *J. curcas* has arisen because it is a relatively "tough" plant. It is tolerant to drought and adapts very well on marginal land, which is not favourable for agricultural use and typically will not be used for crop cultivation (Cox et al., 2006, Glover et al., 2007).

Each *J. curcas* fruit typically bears three seeds (Figure 1.4D). The typical seed mass of *J. curcas* is about 600 mg, with 60–70% being the seed kernel (King et al., 2009). The oil content of *J. curcas* seeds varies from 18-42% (Heller, 1996) but typically lies in the range of 30-35% (Brittaine and Lutaladio, 2010). *J. curcas* has not been domesticated and grows mostly in the wild. From different provenances, the seed yield performance is greatly affected by plant maturity, plantation density, environmental factors and soil fertility. Hence the available yield performance data for this species are limited and somewhat uncertain. Yields between 0.1 and 8 tonnes per hectare have been reported (Brittaine and Lutaladio, 2010, Heller, 1996, Jongschaap et al., 2007). Assuming a productive yearly seed yield of 4.5 tonnes, approximately 1.5 tonnes of oil per hectare could be produced. More yield data are likely to become available over the next few years which will assist the estimation of

the annual oil yield of *J. curcas*. *J. curcas* oil is rich in unsaturated fatty acids, mainly oleate (18:1) and linoleate (18:2), followed by saturated fatty acids palmitate (16:0) and stearate (18:0) (Achten et al., 2008). *J. curcas* oil is suitable to produce biodiesel which meets the EU and USA criteria (discussed further in Chapter 3).

In addition to its potential for biodiesel production, cultivation of *J. curcas* will help to prevent soil erosion (Brittaine and Lutaladio, 2010, Heller, 1996, King et al., 2009). Taken together, *J. curcas* is particularly attractive as a potential biodiesel feedstock.

1.4.3 Developing other by-products from J. curcas

Apart from the seed oil being a good feedstock for biodiesel production, other by-products can be developed from *J. curcas* (Figure 1.5). Using these by-products could not only add to its economic value, but also contribute to a positive energy balance of *J. curcas* cultivation (Achten et al., 2008). Vyas and Singh (2007) suggested that the husk (dried fruit) and hull (testa) can be used to produce bio-gas. The glycerol generated as a by-product of transesterification can be used in industries for various purposes, such as in cosmetics and pharmacy (Pagliaro et al., 2007).

J. curcas seeds are rich in proteins. The composition of *J. curcas* seed meal compares favourably with soybean meal, containing a good balance of essential amino acids except lysine (Makkar et al., 1998a, Makkar et al., 1998b). This highlights the potential of the seed meal for animal feed after oil extraction (King et al., 2009). However, this is currently unavailable due to a range of toxins and anti-nutrients in *J. curcas* seeds (see the next section). Alternatively, the organic-rich seed meal can be used as a fertilizer (Heller, 1996). Some studies reported that application of *J. curcas* seed cake to the soil increases its own seed yield (Ghosh et al., 2007) and the yield of edible crops (Achten et al., 2008).

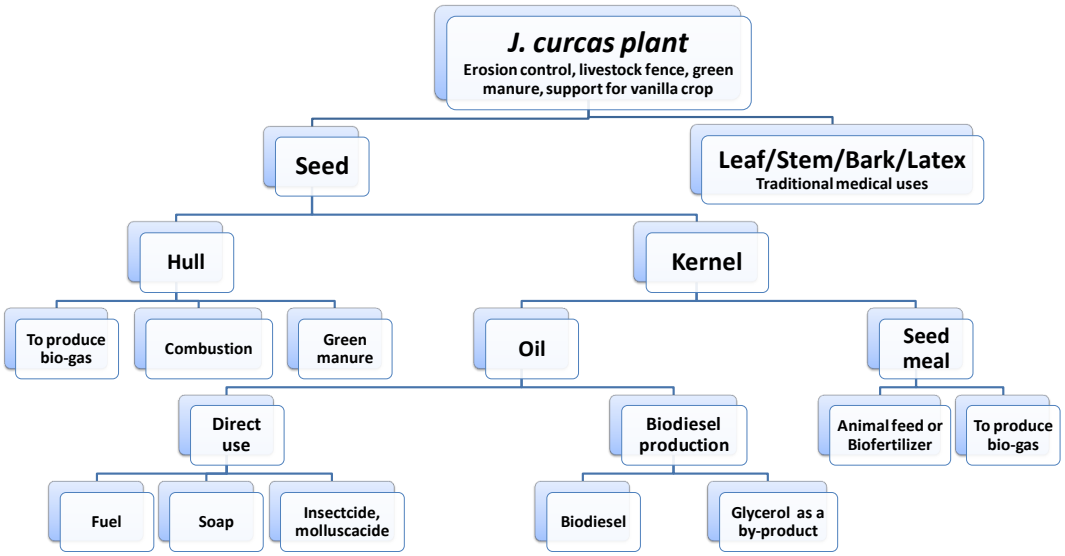


Figure 1.5 Multi-purpose use of *J. curcas* plants. Summarized from Brittaine and Lutaladio (2010), Divakara et al (2010), Heller (1996), Pagliaro (2007) and Vyas and Singh (2007).

1.4.4 Seed toxicity

Most *J. curcas* seeds from present global plantation are known to be non-edible (Heller, 1996). Accidental consumption of non-edible *J. curcas* seeds by humans can cause giddiness, vomiting and diarrhoea. Mortality was observed under force-feeding conditions of non-edible *J. curcas* seeds in laboratory animals (e.g., in chickens, goats and mice) (Makkar and Becker, 2009). Several toxins and anti-nutrients are present in *J. curcas* seeds, including phorbol esters (PEs), curcin, lectin, trypsin inhibitors, phytates, saponins and tannins (Table 1.3). Although seed toxicity will limit the usefulness of these seeds as a source of animal feed, a variety of *J. curcas* bearing edible seeds has been found in Mexico (Makkar et al., 1998a, Makkar et al., 1998b, Makkar and Becker, 2009).

PEs and curcin are considered to be the most toxic compounds in non-edible *J. curcas* seeds (Devappa et al., 2010a). PEs are diterpenoids (will be discussed in details in Chapter 3), which are present in the non-edible variety but lacking in the edible seed variety (Makkar et al., 1998a, Makkar et al., 1998b). Curcin belongs to the ribosome inactivating protein family (will be discussed in details in Chapter 5). It is abundant in the non-edible variety, but it is not known whether or not it is present in the edible variety.

Lectins are a group of sugar-binding proteins which are found in many organisms and involved in a series of biological processes (Elgavish and Shaanan, 1997). When consumed at high levels, lectins threaten the growth and health of animals (Vasconcelos and Oliveira, 2004). Lectin levels in both edible and non-edible *J. curcas* varieties have been found in a similar order of magnitude as in soybean meal (Makkar et al., 2007). Trypsin inhibitors (TI) are protease inhibitors which reduce digestibility in monogastric animals. TI levels are similar in both the non-edible and edible varieties of *J. curcas*, and are similar to that in raw soybean meal (Devappa et al., 2010a).

Phytates are phytic acid [inositol hexakisphosphate (IP6)] or its salt forms, which cannot be digested by nonruminants. Phytates levels in *J. curcas* seeds are high (Devappa et al., 2010a), indicating that phytase as an additive is desirable to improve feed digestibility. Saponins are terpenoids present in many plants, which reduce palatability to livestock. Tannins are phenolic substances which are toxic and antinutritional. Phytates and saponins levels are similar in both *J. curcas* varieties. Tannins level is negligible in both *J. curcas* varieties.

Table 1.3 An overview of seed toxicity in *J. curcas*.

Toxins or anti-nutrients	Adverse effect	Abundance in seeds from two varieties	
		Non-edible variety	Edible variety
Phorbol esters	Interfere with cellular signal pathway; tumor promoter.	Abundant	Undetectable or very low
Curcin	Inhibit protein synthesis in cell free system.	Abundant	Unknown
Lectin	Threaten the growth and health of animals at high levels.	102[1]	51[1]
Trypsin inhibitors	Decreased protein digestibility to monogastrics.	18.4-27.3[2], similar in both varieties.	
Phytates	Reduced intake of mineral ions, phosphorus and dietary proteins to non-ruminants.	7.2-10.1% (w/w[3]), similar in both varieties.	
Saponins	Reduced palatability to livestock	1.8-3.4%[4], similar in both varieties.	
Tannins	Reduced food intake, growth retardation and impaired nutrient absorption	<0.05% (w/w), negligible in both varieties.	

Summarized from Devappa et al (2010a).
1: present as mg of meal that produced hemagglutination per ml of asay medium.
2: mg trypsin inhibited g^{-1}
3: weight/weight base.
4: as diosgenin equivalent.

Comparative feeding studies have been performed using seed meal from both edible and non-edible *J. curcas* varieties. Heat-treated meal of an edible *J. curcas* variety was shown to have no adverse effects on carp and rats (Makkar and Becker, 1999). Conceivably, heating treatment will inactivate curcin, trypsin inhibitors, lectins and increase digestibility of the seed meal (Aderibigbe et al., 1997, Makkar et al., 1998a). PEs, phytates and saponins are not heat labile but the presence of the latter two do not cause poisoning. These feeding studies revealed that PEs are the main toxic agent in *J. curcas* seeds and the main reason why the seed meal is unsuitable as animal feed.

Efforts have been developed to study the detoxification process for *J. curcas* seed meal. Extraction with 90% ethanol, followed by treatment with 0.07% $NaHCO_3$ and heating considerably decreased the PEs content by 97.9% (Martinez-Herrera et al., 2006). Hydrolysis using cellulase and pectinase followed by washing with 65% ethanol eliminate PEs (Xiao et al., 2011). Complete degradation of PEs by solid-state fermentation using the *Pseudomonas aeruginosa* PseA strain was also reported (Joshi et al., 2011). It is not clear if these detoxification treatments are economical at a large scale, and feeding studies would be necessary to confirm that the detoxified

seed meal has no adverse effect on livestock. Deveppa et al (2010b) showed PEs are also biodegradable in soil, and the degraded products are not toxic to snails. This implies *J. curcas* seed meal from the non-edible variety could be used as a fertilizer. Further studies are needed for the thorough evaluation of the eco-toxicity of PEs, such as impact on the groundwater system.

1.5 Targets for the improvement of *J. curcas* as an oilseed crop

Although *J. curcas* has some history of large scale cultivation (Heller, 1996), compared with the other major oilseed crops, the farming of *Jatropha* on an industrial scale can be regarded as a very recent development. Consequently, there are many aspects of *J. curcas* cultivation which could be improved:

1. Seed yield and oil content could be increased.
2. The fatty acid composition could be altered to produce an optimal feedstock for biodiesel production.
3. Synchronized flowering may be desirable to reduce the input of labour in harvesting.
4. Detoxification of *J. curcas* seeds or the seed meal could be performed to improve the use of the meal as animal feed.

Seed yield could be increased through improved cultivation practices, such as irrigation, fertilization management, weeding, pruning, canopy management, and the use of bees as pollinators. Phytohormone treatments may also increase yield; one study reported up to 11-fold seed yield improvement through reducing vegetative growth by applying growth retardant paclobutrazol (Ghosh et al., 2010). Optimization of fatty acid composition and synchronized flowering could be achieved by genetic engineering (King et al., 2009). Detoxification of seed meal by physical and chemical treatments at lab scale have been reported (Joshi et al., 2011, Xiao et al., 2011), but it is not known if economical methods would be available for large scale seed meal production.

As *J. curcas* plantations are currently using wild varieties, perhaps the best opportunity for the improvement of *J. curcas* as an oilseed crop would be the development on new varieties by plant breeding. Plant breeding programs have contributed dramatically to the increase of yields in all major oilseed crops (Buchanan et al., 2000) and oil palm (Soh et al., 2003). A similar programme of plant breeding for *J. curcas* would also be likely to lead to the aforementioned objectives. Conventional and molecular breeding would both be suitable, but both of these approaches need a better understanding of the genetic diversity in *J. curcas*.

1.6 Genetics of *J. curcas*

For *J. curcas* breeding aimed at the improvement of agronomic traits, it is vital to identify varieties with high levels of both phenotypic and genetic diversity. This allows for the preservation of core varieties in germplasm banks (Gepts, 2006), which can be used in genotype versus environment studies, or selecting parental lines for the creation of hybrid plants or mapping populations (Schneider, 2005). To identify genetically distinct accessions, molecular markers are a useful tool. A number of marker systems can be used for genotyping. For *J. curcas*, very few genetic diversity studies were performed before 2008, though in the last three years some have become available. Genetic diversity in *J. curcas* will be discussed in detail based on the present work and analyses by other groups in Chapter 4.

The estimated genome size of *J. curcas* is 410 Mb (Carvalho et al., 2008), about 3.3 times the genome size of *Arabidopsis* (125 Mb) (Kaul, 2000). This relatively small genome of *J. curcas* makes it feasible for genome sequencing and transcriptome sequencing is also possible. In particular, commercialized pyrosequencing such as 454 sequencing (Rothberg and Leamon, 2008) and Illumina sequencing (Morozova and Marra, 2008) have significantly accelerated the process of acquiring ESTs or genome sequences from species of interest. These sequencing techniques, together with the conventional terminator sequencing (Sanger method) have been applied to *J. curcas* research to obtain extensive DNA sequence information (this topic will be discussed in Chapter 4 and 6).

1.7 Aims of this study

1.7.1 Characterization of seed composition and seed oil mobilization

As a potential source of biodiesel, *J. curcas* seed oil content and fatty acid composition is of vital importance because they determine the oil yield and quality of the biodiesel. Analyses were performed in order to establish if differences exist in seed oil quality and quantity in different provenances. To distinguish seed edibility, phorbol ester content was determined. Seed oil mobilization has been studied in model plants such as *Arabidopsis* and also in *J. curcas* recently. To better understand seed oil mobilization, lipids breakdown in germinating *J. curcas* was studied and compared with previous studies.

1.7.2 Characterization of genetic diversity

AFLP (Amplified Fragment Length Polymorphism) analysis was performed to establish the degree of genetic diversity in *J. curcas* from distinct locations. Furthermore, this analysis would allow a comparison of genetic variation and seed composition to establish if there is any correlation.

1.7.3 Characterization of curcin

Although curcin is not thought to be the main toxin in *J. curcas* seeds, it is still interesting to understand if curcin levels vary between what are regarded as edible and non-edible seeds. For a more complete understanding of seed toxicity, curcin levels in the edible and non-edible seeds were therefore analysed.

Curcins, as ribosome-inactivating proteins, may play a role in the defence of the plant against pathogen attack or herbivores (see Chapter 5). Hence, as a crop species, it is also interesting to characterise the spatial and temporal expression profile of these proteins in *J. curcas*, including under conditions of stress. Therefore a time-course analysis was performed.

Chapter 2: Materials and Methods

2.1 Plant methods

2.1.1 Plant material

J. curcas seeds were obtained from 23 geographical locations in Madagascar and 14 locations in Mexico. Seed samples from India (1), Brazil (1), Tanzania (1), Suriname (1) and Ghana (1) were purchased from 5 commercial suppliers (Table 2.1).

2.1.2 Time course of germination and seedling establishment

To perform the time-course analysis plant materials during germination and seedling establishment, and the vegetative tissues were collected as follows (Figure 2.1). Seeds from Tanzania were used for the entire set of experiments in this time-course. To sterilize the seed, a bleach tablet (Advanced Sterilization Products, Irvine, CA, USA) was dissolved with 3 mL concentrated hydrochloric acid in a beaker in an air-tight container to generate chlorine gas, and dry seeds were sterilised overnight. To remove the chlorine gas the seeds were aired in a flowhood on the second day. Seeds were imbibed in sterilized water in a beaker at room temperature for 24 hours. Seeds were germinated on Petri dishes on dampened tissues, in a growth cabinet under 16 hours of light at 25 ºC. Six days after imbibition (DAI), the testa was removed and seeds were sown in soil in the green house. From dry seed to the wizened endosperm stage (WE), endosperm and embryo/seedling materials were separated carefully using forceps, instantly frozen in liquid nitrogen and stored in a -80 ºC freezer. Cotyledon was collected in the cotyledon spread (CS) stage, and cotyledon, true leaf, stem and root were collected in the true leaf (TL) stage. For each sample, 10 plant tissues were pooled together and stored in a -80 ºC freezer as described above.

2.1.3 Stress treatments of young seedlings

J. curcas seeds were sterilized, imbibed, and germinated as described in Section 2.1.2. Three to four week old seedlings were subjected to the stress treatments. For

the mechanical wounding treatment, 50% area of young leaves was crushed with a hemostat, with the vein being avoided. For the NaCl and PEG-6000 (Polyethylene Glycol 6000, Sigma-Aldrich, Dorset, UK) treatment, plants were irrigated with 150 mM NaCl and 40% (w/v) PEG-6000 solutions. A control set-1 was set up in parallel without any treatment. For the MeJA treatment, plants were sprayed with 50 µM MeJA in 0.1% (v/v) methanol. A control set-2 was set up in parallel, sprayed with 0.1% (v/v) methanol. Samples subjected to the above four stress conditions were collected at 15 minutes, 30 minutes, 1 hour, 2 hours, 3 hours, 6 hours, 12 hours and 24 hours after the treatment; samples for control set-1 and set-2 were collected at 15 minutes, 2 hours, 6 hours and 12 hours. For each sample, leaf materials from ten plants were pooled together, immediately frozen in liquid nitrogen and stored at -80 ^{o}C.

Table 2.1 Collection of *J. curcas* seeds used in the present study.

Provenance	Key	Approximate co-ordinates	
		Latitude	Longitude
Madagascan sites			
Trajavona, Ambalavoa	AM1	-21° 51' 29.70"	+46° 54' 30.24"
Mahavanona, Ambalavoa	AM2	-21° 49' 42.36"	+46° 51' 44.16"
Ankaramena, Ambalavoa	AM3	-21° 56' 49.62"	+46° 38' 16.44"
Andalatanosy, Androy	AN1	-24° 40' 01"	+45° 35' 55"
Marovoay, Mahajanga	BO1	-16° 05' 40"	+46° 38' 51"
Ampitolova, Mahajanga	BO2	-15° 39' 16.02"	+46° 22' 46.26
Ambaiboho, Lac Alaotra	LA1	-17° 45'	+48° 30'
Manganoro, Lac Alaotra	LA2	-17° 52'	+48° 23'
Ankasina, Lac Alaotra	LA3	-17° 26'	+48° 37'
Mahasolo, Moyen Ouest	MO1	-19° 06' 04.38"	+46° 21' 09.12"
Ambohikambana, Moyen Ouest	MO2	-19° 04' 27"	+46° 01' 23"
Rivière Sakay, Moyen Ouest	MO3	-19° 06' 32.64"	+46° 24' 15.42
Ambatofolaka, Moyen Ouest	MO4	-18° 58' 33.30"	+46° 32' 08.16"
Amparaky, Moyen Ouest	MO5	-18° 57' 15.42"	+46° 40' 32.88"
Site 1, Antalaha	SA1	N/A	N/A
Site 2, Antalaha	SA2	N/A	N/A
Andasy, Soavina	SO1	-20° 20' 14.10"	+46° 55' 24.00"
Fitamatsina, Soavina	SO2	-20° 21' 01.50"	+46° 55' 58.02"
Befoly, Toliary	TO1	-22° 02'	+43° 46'
Andranovory, Toliary	TO2	-22° 49'	+44° 19'
Lokomby, Vatovavy Fito Vinany	VF1	-22° 11' 30.23"	+47° 44' 55.54"
Kelilalina, Vatovavy Fito Vinany	VF2	-21° 28' 49.97"	+47° 56' 01.29"
Kianjavato, Vatovavy Fito Vinany	VF3	-21° 23' 00.67"	+47° 52' 24.89"
Mexican Sites			
Buena Vista, Puebla	BV	+20° 07' 40"	-97° 27' 15"
Cucuchuchut, Puebla	CCCC	+20° 05' 40"	-97° 35' 49"
Ecatlán (x3 sites), Puebla	ECAT	+20° 03' 09"	-97° 33' 24"
El Naranjo, Puebla	NARAN	N/A	N/A
Tetelilla (x3 sites), Puebla	TET	+20° 04' 41"	-97° 32' 59"
Hueytamalco , Puebla	HUEY	+19° 56' 25"	-97° 17' 17"
Huehuetla , Puebla	HUEH	+20° 06' 12"	-97° 37' 30"
Zozocolco de Hidalgo, Vera Cruz	ZZ	+20° 08' 26"	-97° 34' 34"
Bacalar, Quintana Roo	BAC	+18° 40' 55"	-88° 23' 54"
Rosario Izapa, Chiapas	ROS	+14° 58' 30"	-92° 09' 16"
Commercially supplied seeds (unknown origin)			
Diligent, Tanzania	TANZA	N/A	N/A
Tropilab, Suriname	SUR	N/A	N/A
Brazil	BRA	N/A	N/A
Ghana	GHA	N/A	N/A
India	IND	N/A	N/A

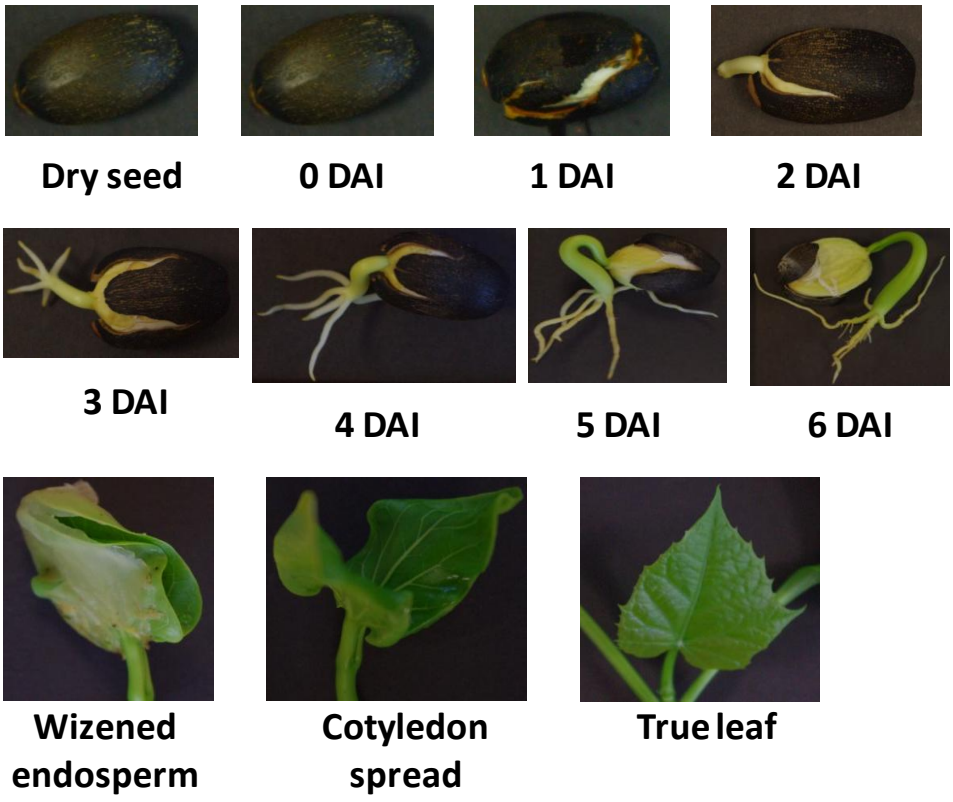


Figure 2.1 Germination and seedling establishment, and vegetative tissues for the time-course analysis. DAI: days after imbibition.

2.2 Biochemical analyses of seed composition

2.2.1 Seed mass, seed mass distribution and kernel percentage

200 *J. curcas* mature dry seeds from 23 provenances in Madagascar (Table 2.1) were weighed individually. Average seed mass and seed mass distribution at 20 mg intervals were calculated accordingly. Testa were removed from 30 seeds from each provenance, and the kernel mass was measured. The average percentage of kernel mass as whole seed mass in different provenances was calculated.

2.2.2 Measurement of oil content and fatty acid composition by FAMEs

Oil analysis by FAMEs (Fatty Acid Methyl Esters) was performed on *J. curcas* mature dry seeds from 23 provenances in Madagascar (Table 2.1). In the preliminary experiment it was observed that seeds with very low mass significantly biased the average oil content measurement, probably due to the kernel being

abnormally shrivelled. The oil content in these wizened seed kernels is often less than 20% as kernel mass (data not shown). Thus to obtain an accurate average oil content, a cut-off was applied for those seeds whose mass lies within the lowest 10% based on the 200 seed mass distribution.

After removal of testa, 30 seeds were ground individually with a mortar and pestle. 10-20 mg portion of the ground material of each seed was transferred into a 2 ml screw-top vial. 25 µl of 25 mg ml^{-1} tripentadecanoic acid (Sigma-Aldrich) was added as the internal standard. 200 µl hexane and 500 µl of 1 M methanolic HCl (Sigma-Aldrich) were added into the vial. To convert fatty acids into their fatty acid methyl esters, transmethylation was performed by heating the samples at 85 °C for 2 hours. After 2 hours, samples were cooled to room temperature, and 250 µl of 0.9% (w/v) KCl was added. 800 µl hexane was then added into the vial followed by brief vortexing. 100 µl FAMEs extract in the supernatant was diluted in 400 µl hexane. FAMEs samples were analyzed by gas chromatography (GC) as described by Larson and Graham (2001) with minor modification. A 1 µl aliquot was injected for FAMEs analysis. Fatty acids were identified and quantified by comparison to a 37 FAMEs mix (Supelco), using Chromquest 2.53 software (Thermo). The oil content (as total fatty acid content) relative to seed mass and composition of each fatty acid was calculated.

A formula has recently been published to estimate the cetane value (CN) from the fatty acid profile of oils, which is reported to give 88% accuracy (Bamgboye and Hansen, 2008). This formula has been applied to the analysis of seed oil samples collected from Madagascar.

2.2.3 Quantification of phorbol esters content

Phorbol esters content was quantified on *J. curcas* mature dry seeds from 23 provenances in Madagascar, 14 provenances in Mexico, 1 provenance each from Suriname and Tanzania (Table 2.1). Testa was removed from mature, dry seeds of *J. curcas*. The seed kernel was ground to a fine powder using a mortar and pestle. 100-200 mg portions of the ground kernels were transferred to a 10 ml conical borosilicate glass centrifuge tube. 50 µg of phorbol 12-myristate-13-acetate (PMA)

(Sigma-Aldrich) was added as an internal standard. Phorbol esters were extracted once with 4 ml of 3/2 hexane / isopropanol for 1 hour with agitation on a rotary platform. The solvent was recovered after centrifugation of the extract at 1,000 *g* for 10 minutes. Two further extractions were performed with 3 ml of hexane/isopropanol and 3 ml of isopropanol. The extracts were combined and the solvent was removed under a stream of nitrogen. The oily residue was dissolved in 2 ml hexane. An equal volume of acetonitrile was added, and the extract was then vortexed. The hexane layer was removed after centrifugation at 1,000 *g* for 10 minutes, and two hexane washes of 2 ml were performed to remove neutral lipids. The washed acetonitrile fraction was filtered through a 0.45 µM syringe filter (Millex®-LH, Millipore, Billerica, MA) and then concentrated to a volume of approximately 200 µl in a Speedvac. 20 µl of the extract was analysed by HPLC using a Phenomenex Luna C18 (2) column (particle size 5 µM, pore size 100 Å, Macclesfield, Cheshire, UK). Elution solvents were (A) water with 0.1 % formic acid and (B) acetonitrile with 0.1 % formic acid. Flow conditions were 5 % solvent B for 0-5 minutes, linear gradient to 80 % B from 5-40 minutes, linear gradient to 100 % B from 40-50 minutes, 100 % solvent B for 50-60 minutes. The column was maintained at 40 °C. Phorbol ester concentrations were calculated from integrated peak areas obtained at 240 nm with reference to the internal standard (PMA) peak area. The detection limit of PE measurement in the current analysis is about 1 µg g^{-1} seed meal.

2.3 AFLP analysis

Seed materials listed in Table 2.1 were used for AFLP analysis. Seeds were germinated in the greenhouse. For each provenance, up to 4 individual *J. curcas* DNA samples were extracted as the DNA template of AFLP (In some provenances the seeds were difficult to germinate thus 1~2 DNA samples were prepared).

The AFLP protocol was modified from Vos et al (1995) by Dr. Andrew King.

2.3.1 DNA extraction and oligonucleotide adapter preparation

Plant genomic DNA was extracted using Qiagen Plant Mini kit (Qiagen, Hilden, Germany). Young leaf (50 mg to 100 mg) material was ground in a 1.5 ml Eppendorf tube with a blue pestle, and processed according to the manufacturers' protocol. DNA quality and concentration were determined on UV spectrophotometer. Good quality DNA should have an A260/280 ratio of 1.7-1.9, and an A260/230 ratio of 1.9-2.1. The A320 should be nil. Significant deviation outside these values indicated contamination, in which case the DNA was further purified by chloroform: isoamyl alcohol extraction (24:1 v/v) followed by ethanol precipitation.

Double-stranded EcoRI, MseI and TaqI adapters were prepared as follows. Single stranded DNA oligonucleotides Eco-Ad1 and Eco-Ad2 were mixed at a concentration of 5 μM each in low TE buffer (10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA). For MseI and TaqI single stranded DNA oligonuleotides Mse-Ad1 (or Taq-Ad1) and Mse-Ad2 (or Taq-Ad2) were mixed at a concentration of 50 μM each, respectively. To anneal, the mixture of two adapters were initially incubated at 94 °C for 4 m 45 s, and then incubated for 15 s stepwise from 94 °C to 20 °C with cooling rate at 0.2 °C per step. The sequences of each pair of oligonucleotides used to create the adapters are listed in Table 2.2.

Table 2.2 Oligonucleotides used as adapters and PCR primers for AFLP analysis

Oligonucleotidename	Sequence
Oligonucleotides used for adaptors	
Mse-Ad1	5’-GACGATGAGTCCTGA-3’
Mse-Ad2	5’-TACTCAGGACTCTA-3’
Taq-Ad1	5’-CGTTCAGGACTCAT-3’
Taq-Ad2	5’-GACGATGAGTCCTGAA-3’
Eco-Ad1	5’-CTCGTAGACTGCGTACC-3’
Eco-Ad2	5’-AATTGGTACGCAGTCTAC-3’
Unlabelled MseI primers	
M-C (preamp.)	5’-GATGAGTCCTGAGTAAC-3’
M-CG	5’-GATGAGTCCTGAGTAACG-3’
M-CAA	5’-GATGAGTCCTGAGTAACAA-3’
M-CAG	5’-GATGAGTCCTGAGTAACAG-3’
M-CTT	5’-GATGAGTCCTGAGTAACTT-3’
Unlabelled TaqI primers	
T-C (preamp.)	5’-GATGAGTCCTGAACGAC-3’
T-CC	5’-GATGAGTCCTGAACGACC-3’
T-CG	5’-GATGAGTCCTGAACGACG-3’
T-CT	5’-GATGAGTCCTGAACGACT-3’
T-CAA	5’-GATGAGTCCTGAACGACAA-3’
T-CAC	5’-GATGAGTCCTGAACGACAC-3’
IR-Dye labelled EcoRI primers	
E-A (preamp.)	5’-GACTGCGTACCAATTCA-3’
E-AAC	5’-GACTGCGTACCAATTCAAC-3’
E-ACA	5’-GACTGCGTACCAATTCACA-3’
E-ACC	5’-GACTGCGTACCAATTCACC-3’
E-AAG	5’-GACTGCGTACCAATTCAAG-3’

2.3.2 Restriction digest and adapter ligation

Restriction digestion-ligation reactions (R/L) were performed using either EcoRI/MseI or EcoRI/TaqI combination. For EcoRI/MseI, a one-step overnight R/L reaction was performed at room temperature in 20 µl reactions with 250 ng genomic DNA, 5 U EcoRI (Fermentas, St. Leon-Rot, Germany), 1 U MseI (NEB, Ipswich, MA), 2 U (NEB) T4 ligase, 5 pmol EcoRI adapter and 50 pmol MseI adapter in NEB T4 ligase buffer (1X) supplemented with an additional 50 mM NaCl. EcoRI/TaqI R/L reactions were performed in two-steps. TaqI digestion was performed in 250 ng genomic DNA in a 15 µl reaction in 1x Fermentas EcoRI buffer with 10U TaqI for 4 hours at 65 °C. After cooling to room temperature, the reaction was supplemented with 7.5 µl of a mixture containing 5 U EcoRI, 0.75 µl 10X Fermentas EcoRI buffer, 3U NEB T4 ligase, 5 pmol EcoRI adapter, 50 pmol TaqI adapter, and 12.5 nmol ATP (pH 7.0). The incubation was continued overnight at room temperature.

2.3.3 Preliminary amplification, selective amplification and gel visualization

The entire set of R/L DNA samples, including 1 negative control (no template control) and 8 positive replicates were randomized for AFLP analysis. The positive replicates were included to verify the reproducibility of the experiment and the randomized samples' order serves to prevent the risk of investigator-associated scoring biases. Pre-amplification and selective amplification reactions were carried out as described previously (Vos et al., 1995) using primers shown in Table 2.2. Pre-amplification products were diluted 100-fold as the DNA template for selective amplification. PCR products of selective amplification were analysed on LI-COR 4300 DNA analyzer (LI-COR Biosciences, Lincoln, NA). 6% polyacrylamide gels (SequaGel XR, National Diagnostics, Atlanta, GA) were prepared according to the manufacturer's protocol. Amplicon sizes were calculated by reference to a 50-700 bp DNA ladder (LI-COR) alongside the gel.

2.3.4 Gel scoring and data analysis

Amplicons between 50-700 bp were scored manually as present ("1") or absent ("0") and inputted into a binary matrix. The percentage of polymorphic bands (PPB) was calculated using POPGENE version 1.32 (Yeh and Boyle, 1997). To evaluate the discriminatory power of each primer combination, polymorphic information content (PIC), marker index (MI) of each AFLP primer set was calculated. PIC values for each locus were calculated using the formula PIC*i* = 2f*i*(1-f*i*), where f*i* is the frequency of the present allele (Roldan-Ruiz et al., 2000). The maximum possible PIC*i* value with this formula for dominant markers is 0.5. The mean PIC values calculated for each primer pair included both monomorphic and polymorphic loci (PIC_{av}) and for polymorphic bands only ($PIC_{av(p)}$). The MI for each primer pair was calculated by multiplying the $PIC_{av(p)}$ value by the number of polymorphic bands (Powell et al., 1996).l

PowerMarker version 3.25 (Liu and Muse, 2005) was used to construct a NJ phylogenetic tree based on Nei's 1973 distance method. Bootstrap value was calculated based on 1000 random re-samplings. Furthermore, I employed Bayesian analysis (character-based method) implemented in STRUCTURE version 2.31 (Pritchard et al., 2000) to delineate population structure. Analysis was carried out for 5 independent runs, each run with burn-in of 100,000, MCMC (Markov Chain Monto Carlo) iterations of 200,000 per chain, and assuming an admixture model and K=1-10. Best number of clusters (No. of K) was identified when LnP (D) values reached a plateau, which means there was negligible increase in LnP (D) in the next K (Evanno et al., 2005). Population membership of STRUCTURE analysis results were graphically displayed using *Distruct* (Rosenberg, 2004).

2.4 Curcin characterization and induction of curcin by stress

2.4.1 Cloning of curcin genes

2.4.1.1 RNA extraction and cDNA synthesis

For RNA extraction, young leaf and dry seed material (about 200-500 mg) was ground with a mortar and pestle in liquid nitrogen, and RNA was extracted using the CTAB method according to Gasic et al (2004). RNA concentration was measured with a Nanodrop spectrophotometer (NanoDrop Technologies, USA). Good quality RNA should have an A260/280 ratio of 1.9-2.2, and an A260/230 ratio of above 2.0. RNA of 5 µg was reverse transcribed to cDNA in a final volume of 20 µL using Oligo (dT) $_{12\text{-}18}$ and SuperScript® II Reverse Transcriptase (Invitrogen, Paisley, UK) according to the manufacturer's protocol. For working stock cDNA was diluted 12-fold in ddH_2O.

2.4.1.2 PCR and PCR product purification

DNA and cDNA were prepared as described in Section 2.3.1 and Section 2.4.1.1. Primers to clone curcin genes (Table 2.3) were designed based on the 454 database previously available in our lab (King et al., 2009). All PCR primers were synthesized by Invitrogen and stock solutions of 100 mM were made by diluting in low TE buffer (10 mM Tris, 1 mM EDTA, pH8.0). For working stock it was further diluted to 5 mM. PCR amplifications were performed in 20 µL volumes. PCR reactions were performed in a Peltier Thermal Cycler DNA Engine (Bio-Rad, Hertford, UK).

PCR was performed using Phusion® High-Fidelity DNA Polymerase (Finnzymes, Espoo, Finland). 1 µL cDNA or gDNA was used as template for each 20 µL reaction containing reaction buffer, dNTPs, forward/reverse primers and proofreading DNA polymerase at concentrations recommended by the manufacturer. Initial denaturing condition was 98 °C for 2 min, followed by 35 cycles of 98 °C for 10 s, 55°C for 30 s and 72 °C for 30 s. The final extension conditions were 72 °C for 3 min followed by a hold temperature of 4 °C.

PCR products were electrophoresed on 0.8% agarose gels. 1 kb or 100 bp DNA ladders (NEB) were used as molecular weight markers. PCR products with the expected size were purified using QIAquick PCR Purification Kit or QIAquick Gel Extraction Kit (Qiagen) as instructed by the manufacturer's protocols.

Table 2.3 PCR and sequencing primers for curcin genes cloning.

Oligonucleotidename	Sequence
Curcin1 F1	5'- TTGCTCAGTTGCTTTCTTTG -3'
Curcin1 R2	5'- GAACTAGTAGAAACTTTATTTG-3'
Curcin3 F3	5'- CATTAATTCCTTCTTGCTCTTGCG -3'
Curcin3 R2	5'- CCCAAGCAAAGCAATACAAA -3'
pJET-F	5'- CTTGGAGCAGGTTCCATTCATTG -3'
pJET-R	5'- CGGTTCCTGATGAGGTGGTTAGC -3'

2.4.1.3 Cloning curcin genes in *E. coli*

Blunt end PCR products were ligated to pJET1.2/blunt cloning vector (Fermentas) according to the manufacturer's instructions. Aliquots of 50 µL chemically competent cells of *E. coli* DH5α strain were defrosted on ice, and 5 µL of the ligation products were added and mixed by flicking. The mixture was left on ice for 30 minutes, heat shocked for 40 seconds at 42 °C and returned to ice for a further 2 minutes. 200 µL sterile LB was added and the cells were incubated at 37 °C, 250 rpm for 1 hour. The transformed cells were plated onto LB agar with 100 µg mL^{-1} ampicillin. Plates were incubated overnight at 37 °C. Single colonies were inoculated into 5 ml LB with 100 µg mL^{-1} ampicillin, and incubated at 37 °C, 250 rpm overnight.

2.4.1.4 Plasmid preparation and sequencing

The overnight *E. coli* cultures were harvested by centrifugation for 10 minutes at 4000 *g*. The plasmids were extracted and purified using the QIAprep Spin Miniprep kit (Qiagen) following the manufacturer's instructions. Plasmid DNA concentration was determined using the Nanodrop spectrophotometer.

DNA sequencing was performed by the DNA Sequencing & Services in University of Dundee (Dundee, UK). Sequencing primers are listed in Table 2.3.

2.4.1.5 *In silico* sequence analyses

Sequence analyses (contigs assembling, sequence alignment) were performed using Vector NTI® Software (Invitrogen). Amino acid sequences were deduced using the Translate tool integrated in ExPASy (www.expasy.org). Intron prediction was performed in the NetGene2 server (http://www.cbs.dtu.dk/services/NetGene2/).

2.4.2 Characterization of seed curcin proteins

2.4.2.1 *In silico* protein sequence analyses

Compute pI/Mw and SignalP tool integrated in ExPASy (www.expasy.org) were used to predict the pI/Mw and the signal peptide of the four curcin protein sequences.

2.4.2.2 Protein extraction

Proteins were extracted as reported previously with minor modifications (Lin et al., 2003a). Seed kernels were ground to fine powder using a mortar and pestle. The ground kernel was transferred into a 2.0 ml Eppendorf tube and extracted with hexane for three times to remove the oil. Residual hexane was removed using a SpeedVac. 100 mg of the defatted ground kernel was homogenized in 1 ml extraction buffer (50 mM Na_3PO_4, 0.2 M NaCl, pH 7.2). The supernatant was recovered after centrifugation at 13,000 g for 5 minutes. This extraction procedure was repeated twice and the recovered supernatants were pooled. Extracts were also obtained from tegmen, embryo and endosperm using the same method. Protein concentration of the pooled extract was determined according to Bradford (1976) using bovine serum albumin as a standard.

2.4.2.3 SDS-PAGE and Coomassie staining

Proteins (20 µg) were separated on 14% w/v SDS-PAGE gel (Laemmli, 1970) using Mini-PROTEAN®3 Cell (Bio-Rad) according to the manufacturer's instructions. Gels were stained with a colloidial Coomassie Blue stain (0.1% Coomassie Blue G-

250, 17% $(NH_4)_2SO_4$, 34% methanol, 3% H_3PO_4). Protein markers (Unstained protein molecular weight marker, Fermentas) were run on each gel for molecular weight determination.

2.4.2.4 Development of the anti-curcin antibody

A custom polyclonal antibody was raised against keyhole limpet hemocyanin fusion with the peptide corresponding to amino acids 187-201 (H_2N-LEKSSKPADIAKPLVC-COOH) of curcin1 in rabbit (Eurogentec, Seraing, Belgium). It was expected that this anti-curcin antibody would also cross-react with other curcin proteins due to the high similarity in amino acids 187-201 in the four curcins.

2.4.2.5 Heterologous expression of curcin proteins in *E. coli*

To test the specificity of the seed curcin antibody, the DNA sequences corresponding to the full-length coding region of the four curcin proteins were amplified from *J. curcas* seed cDNA using Phusion® High-Fidelity DNA Polymerase using primers listed in Table 2.4. The underlined sequences represent NdeI and BamHI (NEB) restrictions sites respectively. The PCR product was then restriction digested and ligated into pET28a vector (EMD Biosciences, Darmstadt, Germany) to create an N-terminal hexa-histidine tagged construct, which was transformed into *E. coli* DH5α strain. Cloning and sequencing were performed according to Section 2.4.1.3 and 2.4.1.4, except that kanamycin was used for selection. Plasmids containing the positive constructs of the four curcin genes were transformed into Rosetta™ 2 (DE3) cells (EMD Biosciences). The empty vector of pET28a (EMD Biosciences) was also transformed in parallel as the negative control. The transformed *E. coli* strain was grown at 37 ºC, 250 rpm in LB medium supplemented with 34 μg ml^{-1} chloramphenicol and 50 μg ml^{-1} kanamycin to an OD 600 nm of 0.6 prior to induction with 1 mM IPTG. The culture temperature was then reduced to 28 ºC, and 100 ml of cells were harvested 6 hours post-induction by centrifugation and resuspended in 2 ml 1X PBS buffer (Phosphate Buffered Saline, pH 7.4). Sonication was performed in salt/ice bath with an S-4000 Sonicator (Misonix) at 70 % amplitude for intermittent 6 × 10 s cycles. The crude lysates were aliquoted, of which 100 μl from each sample was mixed with 400μl 1X SDS buffer as the crude

extract samples. To prepare for the supernatant extract samples cell lysates were centrifuged at 20,000 *g*, 4 °C for 30 minutes to remove the cell debris. The same dilution was made in 1X SDS buffer as the crude extracts. For SDS-PAGE visualization 10 µl of the samples were boiled at 95 °C for 5 minutes before loading.

Table 2.4 PCR and sequencing primers for curcin expression in *E. coli.*

Oligonucleotide name	Sequence
E.Coli C1-F	5'- TTCATATGAAAGGTGGAAAGATGAATCTCT -3'
E.Coli C1-R	5'- TTGGATCCTCAGACTTTGTAATTGACTGCATTCA -3'
E.Coli C2-F	5'- TTCATATGAAAGGTGGAAAGATGAACCTC -3'
E.Coli C2-R	5'- TTGGATCCTTAAAGCCATGGCAGCCAC -3'
E.Coli C3-F	5'- TTCATATGAAACGAGGAAACACGAAGC -3'
E.Coli C3-R	5'- TTGGATCCTTAGCTGTTGATAATAATTTCTTCCGTG -3'
E.Coli C4-F	5'- TTCATATGAAAGGTGGAAACATGAAGCT -3'
E.Coli C4-R	5'- TTGGATCCTTAAAGCAATGGCAGCCACTT -3'

2.4.2.6 Western blot

For Western blot analysis proteins samples were separated on SDS-PAGE (Section 2.4.2.3) and transferred onto a nitrocellulose membrane (Hybond ECL, GE Healthcare, Buckinghamshire, UK) by electroblotting (Bio-Rad). BenchMark™ Pre-Stained Protein Ladder (Invitrogen) was used for molecular weight determination. Ponceau S staining was used to assess equal loading. The anti-curcin blots were developed using the ECL™ Western Blotting Detection Reagents (GE Healthcare), with a polyclonal goat anti-rabbit IgG-HRP conjugate being used as a secondary antibody (Alpha Diagnostic, San Antonio, Texas). The monoclonal anti-polyhistidine antibody was purchased from Sigma-Aldrich. A 4-chloro-1-naphthol tablet (Sigma-Aldrich) was used as the substrate for visualizing the blots following the manufacturer's instructions.

2.4.2.7 Peptide sequencing by MALDI-TOF

Protein identification was performed by the Technology Facility in the Department of Biology (University of York). For peptide sequencing, proteins bands were excised and in-gel tryptic digestion and MALDI-TOF analyses were performed as described previously (Thompson et al., 2010). Tandem mass spectral data were submitted for database searching using a locally-running copy of the Mascot program (Matrix Science Ltd., version 2.1), through the Bruker BioTools interface (version 3.2). The spectral data were searched against a transcriptome database of developing *J. curcas* seeds (King et al., 2010).

2.4.3 Time-course analyses of curcin during seed germination and seedling establishment, and in the vegetative tissues

2.4.3.1 Quantitative real-time PCR (qRT-PCR) analysis of the gene expression of the four curcins

Plant materials were obtained as described in Section 2.1.2. RNA extraction and cDNA synthesized were carried out as described in Section 2.4.1.1 with two minor modifications. To eliminate the DNA, RNA samples were treated with DNase and precipitated in ethanol, and random hexamers (Invitrogen) were used as primers for cDNA synthesis. qRT-PCR primers for the four curcin genes were designed based on their full length sequences; primers for the housekeeping genes *ELONGATION FACTOR1α*, *ACTIN* and *18S* genes in *J. curcas* were previously designed by Dr. Andrew King based on the transcriptome database of developing *J. curcas* seeds (King et al., 2010). Since the four curcin genes share a very high level of similarity, primer pairs for qRT-PCR were tested for their specificity. Gradient PCR (54 ºC to 64 ºC at 11 intervals as the annealing temperature) was performed using plasmid DNA containing the four curcin genes. Primers were considered to be specific to a certain curcin gene if they only amplified the fragment from the target curcin containing plasmid DNA (see Appendix 5). The optimal annealing temperature (T_a) was selected, and used during the qRT-PCR reaction (Table 2.5).

qRT-PCR was performed in MyiQ Real-Time PCR detection system (Bio-Rad) using iQ SYBR Green Supermix (Bio-Rad) according to the manufacturer's instructions. A dilution series of the plasmid DNA corresponding to the gene were

used to build the standard curve. Four replicates of each sample were used in the analysis. In the preliminary study I tested *ELONGATION FACTOR1α*, *ACTIN* and *18S* genes in *J. curcas*. The result shows that *ELONGATION FACTOR1α* and *ACTIN* exhibit a similar pattern of expression (see Appendix 6). Thus *ACTIN* was chosen as the housekeeping gene in this analysis. The expression level of *ACTIN* was used for normalization.

Table 2.5 Primers for qRT-PCR determination of gene expression in *J. curcas*.

Oligonucleotide name	Sequence	*T*a
C1-F3	5'- GCCAAAGTCATAAATGTAGC -3'	58.3 ºC
C1-R2	5'-ATATGTAATTATCTAATGCCTG -3'	
C2-F1	5'- CTTATCAAATACAGAAATCTGTTG -3'	60.0 ºC
C2-R1	5'- CATTCAACAAGACTCCCATGAG -3'	
C3-F1	5'- GTAGCCAAAGTCATAAATTCA -3'	58.3 ºC
C3-R1	5'- GCATAGTTACCAGTAAATGATAATTC -3'	
C4-F	5'- CCATTACTTATATTCCTGATGAG -3'	60.0 ºC
C4-R	5'- CTGATACTTCTATATCCCCAGC -3'	
Isocitrate_lyase-F	5'- ATGGCCTCATCTTTCTCTGTTC -3'	58.3 ºC
Isocitrate_lyase-R	5'- TGGTCTTCGGGTTAGCTTGA -3'	
Malate_synthase-F	5'- ATGGTAGCGGTTGGTCAATAT -3'	58.3 ºC
Malate_synthase-R	5'- ATCATACCGGCCACGAATCT -3'	
EF1α-F	5'- ACCACTGGTGGTTTGAAGCTGG -3'	58.3 ºC
EF1α-R	5'- CAGGGTTGTAACCCACCTTCTTCA -3'	
18S-F	5'- GCCCGTTGCTCTGATGAT -3'	58.3 ºC
18S-R	5'- GGATGTGGTAGCCGTTTCT -3'	
Actin-F	5'- TGCCATCCAGGCCGTTCTATCT -3'	58.3 ºC
Actin-R	5'- GGAGGATAGCATGTGGAAGAGCG-3'	
LOX2-F	5'- ATAATGCCAACTGAAGA -3'	53.3 ºC
LOX2-R	5'- TAGATAGAACATCCAAAAC -3'	
OPR3-F	5'- TTTCTAGTGGTGGGTTC -3'	53.3 ºC
OPR3-R	5'- ATACTTATTCAAGGGTGC -3'	

The optimal *T*a was determined by gradient PCR.

2.4.3.2 Determination of the protein levels of the four curcins

In order to determine the protein levels of curcin against the total soluble protein, the following extraction method was used. Plant materials were ground with a mortar and pestle in liquid nitrogen and then homogenized with extraction buffer [100 mM Tris/HCl, pH 7.5, 150 mM NaCl, 5 mM EDTA, 2 mM DTT, 0.1% (v/v) Triton X-

100, 10% (v/v) glycerol, and protease inhibitors (plant cocktail, Sigma-Aldrich). Homogenates were centrifuged at 20,000 *g*, 4 °C for 30 minutes, and protein concentration was measured as described by Bradford in Section 2.4.2.2. SDS-PAGE and Western blot were conducted according to Section 2.4.2.3 and 2.4.2.6.

2.4.4 Induction of curcin under stress conditions and JA levels determination

Three to four week old young *J. curcas* seedlings were treated with various stresses including mechanic wounding, NaCl, PEG-6000 and MeJA, and materials were collected according to Section 2.1.3. The following analyses were performed.

2.4.4.1 Curcin gene expression upon response to stress by qRT-PCR analysis

RNA extraction and cDNA synthesis was carried out as described in Section 2.4.3.1, as well as the qRT-PCR analysis. Partial sequences of two JA biosynthesis genes, *LOX2* and *OPR3* in *J. curcas* were identified by searching the transcriptome database of developing *J. curcas* seeds (King et al., 2010). qRT-PCR primers were designed accordingly (Table 2.5). The gene expression levels of *CURCIN1-4*, *LOX2* and *OPR3* were determined by qRT-PCR analysis as described in Section 2.4.3.1.

2.4.4.2 Protein levels of curcin under stress conditions

Total soluble proteins extraction was performed according to Section 2.4.3.2. Protein quantification was carried out according to Section 2.4.2.2. SDS-PAGE and Western blot were performed as described in Section 2.4.2.3 and Section 2.4.2.6, respectively.

2.4.4.3 Oxylipin Analysis

Two hundred micrograms of leaf material was used for each sample. The oxylipin extraction and LC-MS analysis were performed according to Dave et al (2011). JA and OPDA levels were determined. Two measurements were performed under the wounding treatment, and only one measurement was taken under the NaCl and PEG-6000 treatments since no obvious change in JA and OPDA levels was observed.

2.5 Characterization of oil mobilization during seed germination and seedling establishment

2.5.1 Measurement of water, oil content and fatty acid composition

Material of the time-course analysis during seed germination and seedling establishment was described in Section 2.1.2. Material was weighed to obtain the fresh weight (FW) and dry weight (DW) after freeze-dried. Water content was calculated as (FW-DW)/FW %. Oil content was analyzed on a GC machine by FAMEs method according to Section 2.2.2.

2.5.2 Gene expression of MALATE SYNTHASE and ISOCITRATE LYASE

RNA extraction, cDNA synthesis was carried out according to Section 2.4.3.1, as well as the qRT-PCR analysis. Partial sequences of two genes in the glyoxylate cycle, *MALATE SYNTHASE* (*MLS*) and *ISOCITRATE LYASE* (*ICL*) in *J. curcas* were identified by searching the transcriptome database of developing *J. curcas* seeds (King et al., 2010). qPCR primers were designed accordingly (Table 2.5). The gene expression levels of *MLS* and *ICL* were determined by qRT-PCR analysis as described in Section 2.4.3.1.

Chapter 3: Characterization of oil and phorbol ester contents of *J. curcas* seeds

3.1 Introduction

3.1.1 Oil as the main reserve in J. curcas seeds

J. curcas, as a perennial plant which can be used as the feedstock for biodiesel production, has received a surge of interest recently. The oil content of *J. curcas* seeds varies from 18-42% (Heller, 1996) but primarily lies in the range of 30-35% (Brittaine and Lutaladio, 2010). Limited yield data have suggested that mature *J. curcas* plants can produce 1.5 tonnes of crude oil per hectare (Foidl et al., 1996), which is suitable to be transesterified into biodiesel (King et al., 2009). As a potential source of biodiesel production, the oil content and fatty acid (FA) composition of *J. curcas* seeds are of vital importance. The oil content of *J. curcas* seeds, together with the seed yield, remarkably influences its economic value, especially for large scale agricultural production. The FA composition of crude oil (e.g. *J. curcas* oil) determines a number of important properties of biodiesel, such as cetane number (CN), viscosity, cold-flow and oxidation stability (Knothe, 2008). CN measures the delay in the combustion of the fuel from ignition, and high CN indicates short ignition delay in the engine. An appropriate viscosity is required for proper engine performance, influencing flow of the fuel through pipelines and injector nozzles, and atomization of the fuel in the cylinder. Cold-flow is a measurement of the performance of the fuel under cold temperature conditions. Proper oxidation stability is essential for biodiesel during storage and handling, because oxidization leads to the formation of peroxides, acids, gums, and deposits, which would be problematic to the engine performance.

Seed germination is a crucial stage of central biological importance. During germination plants achieve the transition from heterotrophic to photoautotrophic growth (Eastmond and Graham, 2001). This process depends on the consumption of the storage reserves in the seed, which include lipids, starch and proteins. Like many plant species such as sunflower, oilseed rape and soybean, *J. curcas* seeds use oil as the main storage reserve. In these oil rich seeds, oil must be mobilized to provide the carbon source and energy for germination. Storage lipid mobilization has been well

studied in the model plant *Arabidopsis* [reviewed by Graham (2008)]. Oil mobilization commences upon germination, initiated by the hydrolysis of triacylglycerol (TAG) in the oil body (No. 1, Figure 3.1). In the glyoxysomes, acyl-CoA esterified FAs are broken down to acetyl-CoA via β-oxidation (No. 2), which serves as the substrate for the glyoxylate cycle (shown in the green box). Two genes encode specific enzymes to the glyoxylate cycle: malate synthase (MLS) and isocitrate lyase (ICL) (No. 5 & 6, respectively). The glyoxylate cycle produces 4-carbon compounds for subsequent gluconeogenesis. Hexoses produced via the gluconeogenesis (No. 9) consequently fuel the process of seedling establishment.

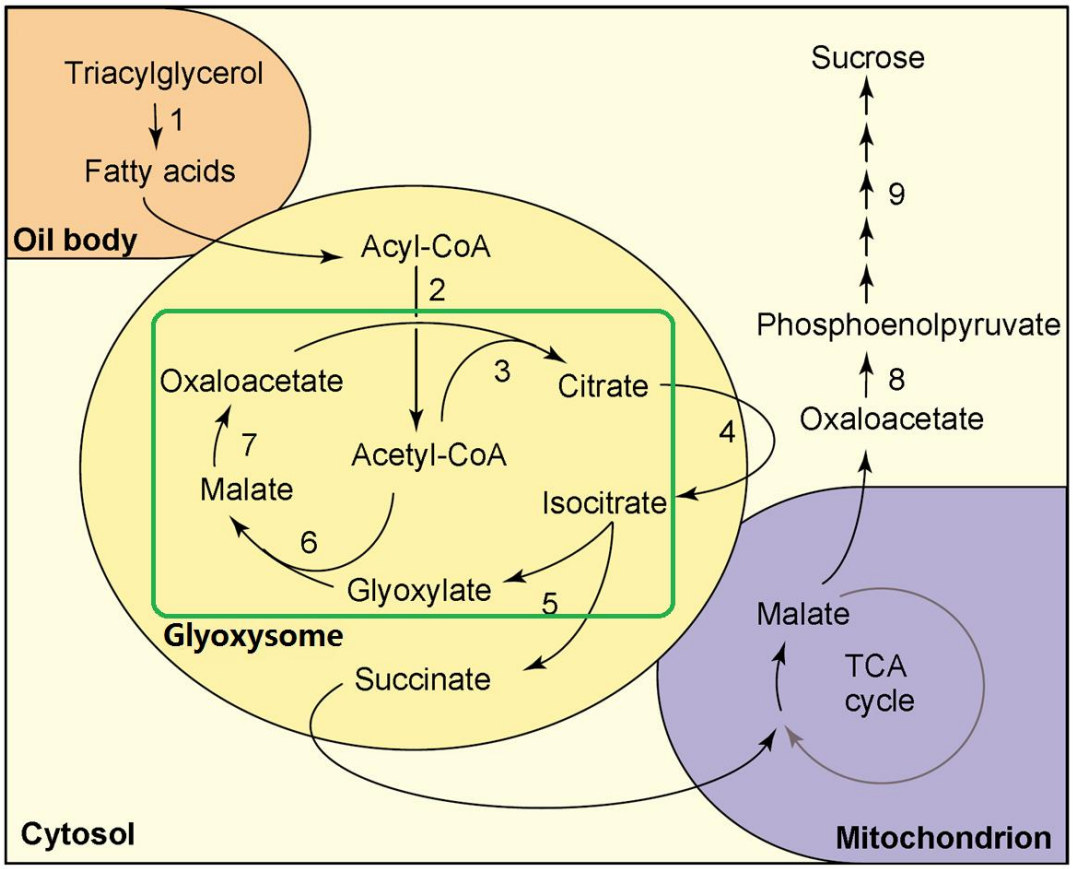


Figure 3.1 The schematic of oil mobilization during seed germination and post-germinative growth. Figure taken from Eastmond and Graham (2001). The green box shows the glyoxylate cycle. Numbers represent metabolic reactions and enzymes: 1, triacylglycerol lipase; 2, fatty acid β-oxidation; 3, citrate synthase; 4, aconitase; 5, isocitrate lyase; 6, malate synthase; 7, malate dehydrogenase; 8, phosphoenolpyruvate carboxykinase; 9, gluconeogenesis.

3.1.2 Phorbol esters as the main toxic agent in J. curcas seeds

Apart from its potential use as biodiesel feedstock, it is recognized that other by-products could also be developed to increase the economic value of *J. curcas* seeds. In particular, the high level of proteins (53-63%) makes *J. curcas* a very good resource for animal feed or bio-fertilizer (Aderibigbe et al., 1997). However, a few toxins or anti-nutrients exist which hamper the utilization of *J. curcas* seed meal as animal feed. Among these toxins or anti-nutrients, phorbol esters (PEs) are the main toxic agents (Devappa et al., 2010a, King et al., 2009). PEs are diterpenoids, and to date six PEs have been identified by NMR (Figure 3.2) (Haas et al., 2002). The phorbol ester biosynthesis pathway is still unclear, but is proposed to involve three steps. First geranylgeranyl pyrophosphate is converted to a cyclic diterpene via the action of a terpene synthase; P450 oxygenases catalyze redox reactions to form phorbol; and acyltransferases transfer acyl groups onto phorbol to form PEs. PEs are present in the seed, leaf, stem, root, flower and bark but not in the latex in *J. curcas* (Makkar and Becker, 2009). PEs are purgative, skin-irritant and, of more concern, are tumour-promoters. PEs act as a structural analogue of diacylglycerol, an activator of protein kinase C (PKC) (Zhang et al., 1995). Unlike diacylglycerol, which has a short biological half-life, the activation of PKC by PEs has a prolonged effect which leads to a number of biological activities, including the triggering of cell proliferation (Goel et al., 2007).

Despite the fact that seeds from most *J. curcas* provenances contain PEs and are therefore non-edible, edible seeds lacking phorbol esters exist in Mexico (Makkar et al., 1998a, Makkar et al., 1998b). These edible seeds are consumed by the indigenous people after cooking (Schmook and Serralta-Peraza, 1997 and J. Cuevas, personal observation). Feeding experiments using heat-treated meal of an edible *J. curcas* variety have shown no adverse effects in carp and rats (Makkar and Becker, 1999).

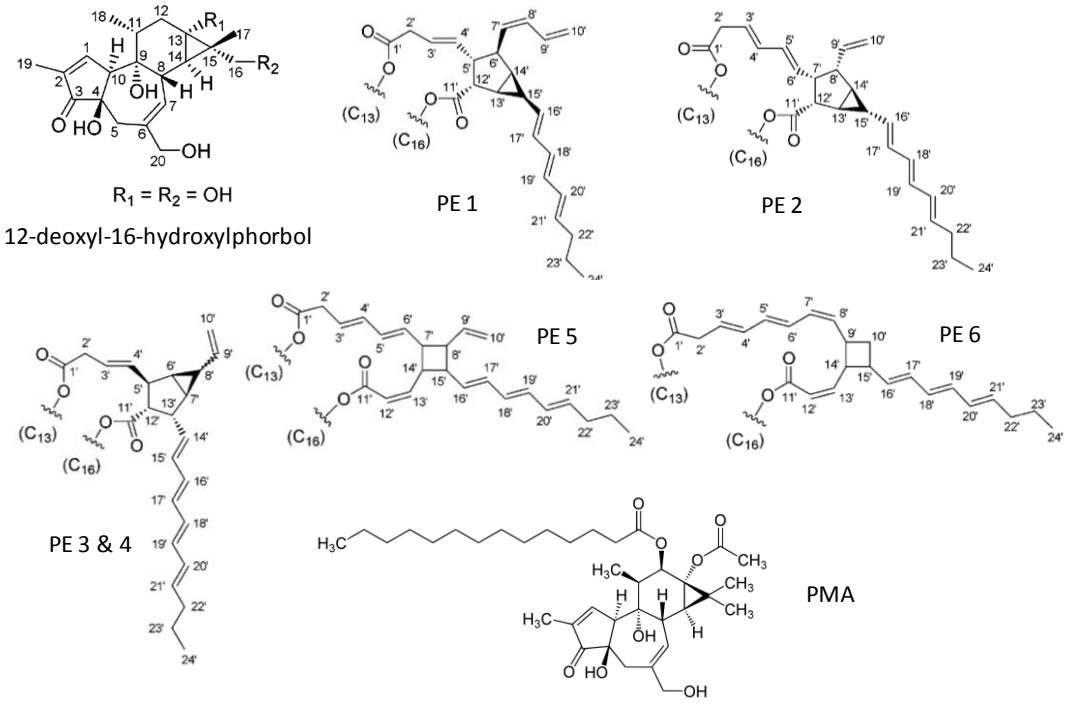


Figure 3.2 Structures of 12-deoxy-16-hydroxylphorbol, six phorbol esters (PEs) in *J. curcas* and PMA (phorbol 12-myristate 13-acetate). Figure adapted from Hass et al (2002). The 12-deoxy-16-hydroxylphorbol is esterified at C-13 (R1) and C-16 (R2) positions to form PEs, and is therefore common to all PEs from *J. curcas*. PMA was used as the internal standard to determine PEs concentration. It has a similar ring structure to 12-deoxy-16-hydroxylphorbol.

3.1.3 Aims

The aims of this study are to:

1. Characterize *J. curcas* seed mass, oil content and FA composition in various seed batches from Madagascar. Measurements of seed oil content and FA composition would establish if differences exist in seed oil quantity and quality, and furthermore if the differences are related to seed mass or environmental conditions.

2. Analyze the phorbol esters content in *J. curcas* seeds from the samples used for the AFLP analysis. This would provide the biochemical basis to distinguish the edible and non-edible seed varieties.

3. Gain a better understanding of lipid mobilization in *J. curcas* endosperm and embryo by studying oil breakdown in these tissues during the germination process.

3.2 Characterization of *J. curcas* seeds

3.2.1 Variation of seed mass, kernel percentage and seed oil content

To characterize *J. curcas* seeds, first seed mass was determined. 200 seeds from each of the 23 provenances in Madagascar were weighted individually. The average seed mass of 23 provenances ranges from 545 to 742 mg, and the average seed mass across 23 provenances is 634 mg (Figure 3.3A). The measurements also revealed patterns of the seed mass frequency in all provenances. Most of the seed mass distributions are in negatively skewed patterns, such as provenance VF3 (Figure 3.4A); and also a few are close to normal distribution, such as provenance BO1 (Figure 3.4B). The average percentage of kernel mass as whole seed mass in different provenances is from 54.0% to 66.3% (data not shown). The average seed mass, kernel percentage, and average kernel mass are presented in Appendix 3.

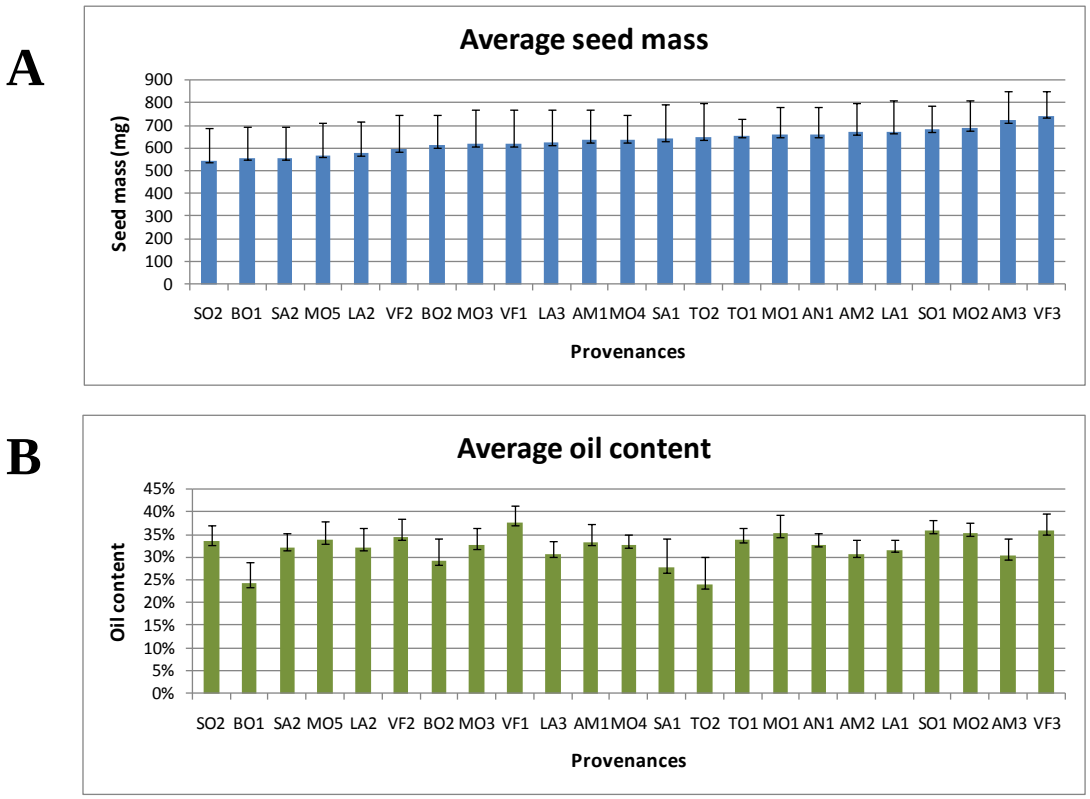


Figure 3.3 Average seed mass (A) and oil content (B) in 23 provenances from Madagascar. Average oil content is represented by the measurement of fatty acids. Provenance codes are given in Table 2.1 in Chapter 2. Error bars represent standard deviation (+) and standard error of the mean (-) of biological replicates.

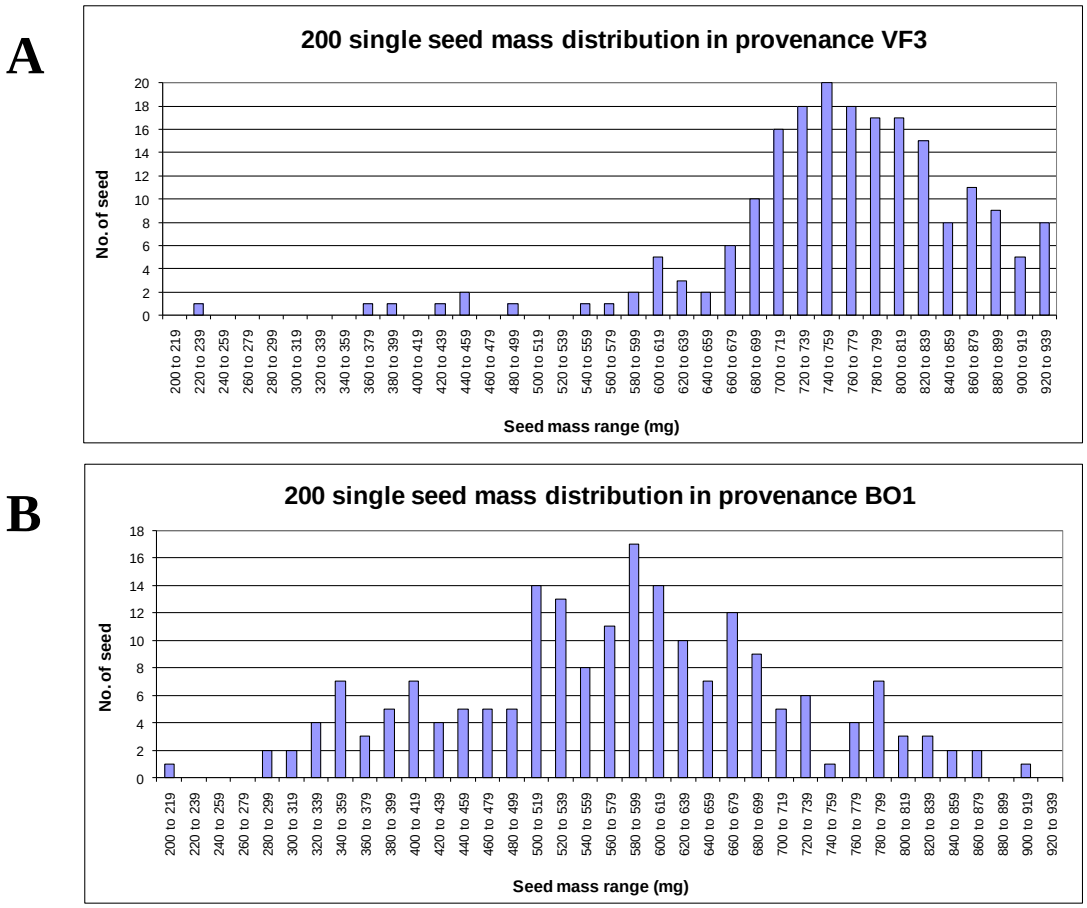


Figure 3.4 Seed mass distribution from two provenances as representative to show negatively skewed (A) and normal distribution (B) pattern.

To compare the oil content in different provenances of *J. curcas* seeds from Madagascar, seeds total FAs were extracted, converted to FA methyl esters (FAMEs) and measured by gas chromatography (GC). FAMEs analyses were performed on 30 individual seeds from each provenance. Since about 95% FAs in oil seeds exist in the form of triacylglycerol (Buchanan et al., 2000), here the oil content in *J. curcas* seeds was represented by the percentage of total FAs, The average oil content in the 23 provenances from Madagascar spans from 24.0% to 37.6% (Figure 3.3B). The average oil content was compared with the seed mass, kernel mass and kernel percentage, but no correlation was observed (comparison in Figure 3.3). Furthermore, altitude and soil composition (nitrogen content, organic matter, potassium, phosphorous, cation exchange capacity, pH, calcium, boron, manganese, copper, magnesium, sulphur, zinc, molybdenum, iron, sodium, lime requirement) data were obtained of the sites where the seed samples were collected from. Data on the mean annual temperature at the different sites were not available. However, the

altitude of each site was available. These data were used as a proxy of the temperature as they are theoretically inversely proportional. Principle component analyses (PCA) were conducted and determined that these environmental factors have no observed impact on the seed oil content (Appendix 1).

3.2.2 Oleate to linoleate ratio shows high variation

In addition to the oil content, the FA composition is also of interest because it determines the oil properties to be used as a feedstock for biodiesel production. Thus, oil content analyses were performed on seeds from the Madagascan provenances, and revealed that oil is composed of approximately 25% saturated and 75% unsaturated FAs (Table 3.1). Palmitate (16:0) and stearate (18:0) are the most abundant saturated FAs, whereas the unsaturated FAs are dominated by oleate (18:1) and linoleate (18:2). There is a significant variation of oleate and linoleate content among the different provenances. The FA composition of the 23 seed batches was further compared with site altitude and soil composition data. It is interesting that the oleate content is inversely proportional to the elevation of the provenance, while the linoleate content increases with the elevation (Figure 3.5). Both correlations are statistically significant ($p<0.05$). Assuming that temperature decreases with the increase of elevation, the result shows that temperature plays an important role in determining the ratio of oleate/linoleate in *J. curcas*. To assess the suitability of *J. curcas* oil as biodiesel feedstock, cetane number (CN) was calculated based on the FAs composition (Table 3.1). The calculated CN value ranges from 51 to 55.

Table 3.1 Fatty acid composition of *J. curcas* seeds in 23 Madagascan provenances.

Provenance	Abbreviation	Elevation	% fatty acid composition								Calculated
		(m)	16:0	16:1	18:0	18:1	18:2	18:3	20:0	Others	CN
Marovoay, Mahajanga	BO1	5	14.3	0.7	7.5	51.7	25.3	0.3	0.3	<0.1	55 (C)
Ampitolova, Mahajanga	BO2	26	14.8	0.8	7.6	48.5	27.8	0.2	0.2	<0.1	55 (C)
Site 1, Antalaha	SA1	ND	14.1	0.7	7.9	49.1	27.7	0.3	0.3	<0.1	55 (C)
Andranovory, Toliary	TO2	ND	14.5	0.9	6.8	45.4	31.9	0.2	0.2	<0.1	54 (C)
Lokomby, Vatovavy Fito Vinany	VF1	20	14.7	0.8	7.5	43.5	33.0	0.2	0.2	<0.1	54 (C)
Site 2, Antalaha	SA2	ND	15.4	0.8	6.9	43.3	33.1	0.2	0.2	<0.1	54 (C)
Befoly, Toliary	TO1	460	14.5	0.9	6.6	42.3	35.3	0.2	0.2	<0.1	54 (C)
Rivière Sakay, Moyen Ouest	MO3	852	16.0	1.1	6.2	37.2	39.1	0.3	0.2	<0.1	53 (C)
Kianjavato, Vatovavy Fito Vinany	VF3	61	15.5	0.9	6.0	37.4	39.8	0.2	0.2	<0.1	52 (C)
Ankaramena, Ambalavoa	AM3	776	14.2	0.9	5.9	39.8	38.8	0.2	0.2	<0.1	52 (C)
Ambaiboho, Lac Alaotra	LA1	754	15.3	1.1	5.6	37.1	40.4	0.3	0.2	<0.1	52 (C)
Andalatanosy, Androy	AN1	460	14.1	0.9	6.0	38.8	39.8	0.2	0.2	<0.1	52 (C)
Ambohikambana, Moyen Ouest	MO2	903	14.9	1.0	5.5	38.1	40.1	0.2	0.2	<0.1	52 (C)
Amparaky, Moyen Ouest	MO5	1145	13.4	0.8	6.6	38.6	40.2	0.2	0.2	<0.1	52 (C)
Mahasolo, Moyen Ouest	MO1	903	14.1	0.9	5.9	37.9	40.7	0.2	0.2	<0.1	52 (C)
Andasy, Soavina	SO1	1084	15.4	1.0	5.4	36.2	41.5	0.2	0.2	<0.1	52 (C)
Manganoro, Lac Alaotra	LA2	772	14.4	0.8	6.9	34.6	42.8	0.2	0.2	<0.1	52 (C)
Kelilalina, Vatovavy Fito Vinany	VF2	620	12.7	0.6	6.1	39.7	40.4	0.2	0.2	<0.1	51 (C)
Ambatofolaka, Moyen Ouest	MO4	974	14.2	0.8	5.8	37.0	41.8	0.2	0.2	<0.1	51 (C)
Ankasina, Lac Alaotra	LA3	833	15.2	1.0	5.3	35.2	42.8	0.2	0.2	<0.1	51 (C)
Trajavona, Ambalavoa	AM1	957	15.0	1.0	5.3	34.9	43.3	0.2	0.2	<0.1	51 (C)
Fitamatsina, Soavina	SO2	1104	15.0	1.0	5.3	35.0	43.3	0.3	0.2	<0.1	51 (C)
Mahavanona, Ambalavoa	AM2	964	15.5	1.1	5.3	33.8	43.9	0.2	0.2	<0.1	51 (C)

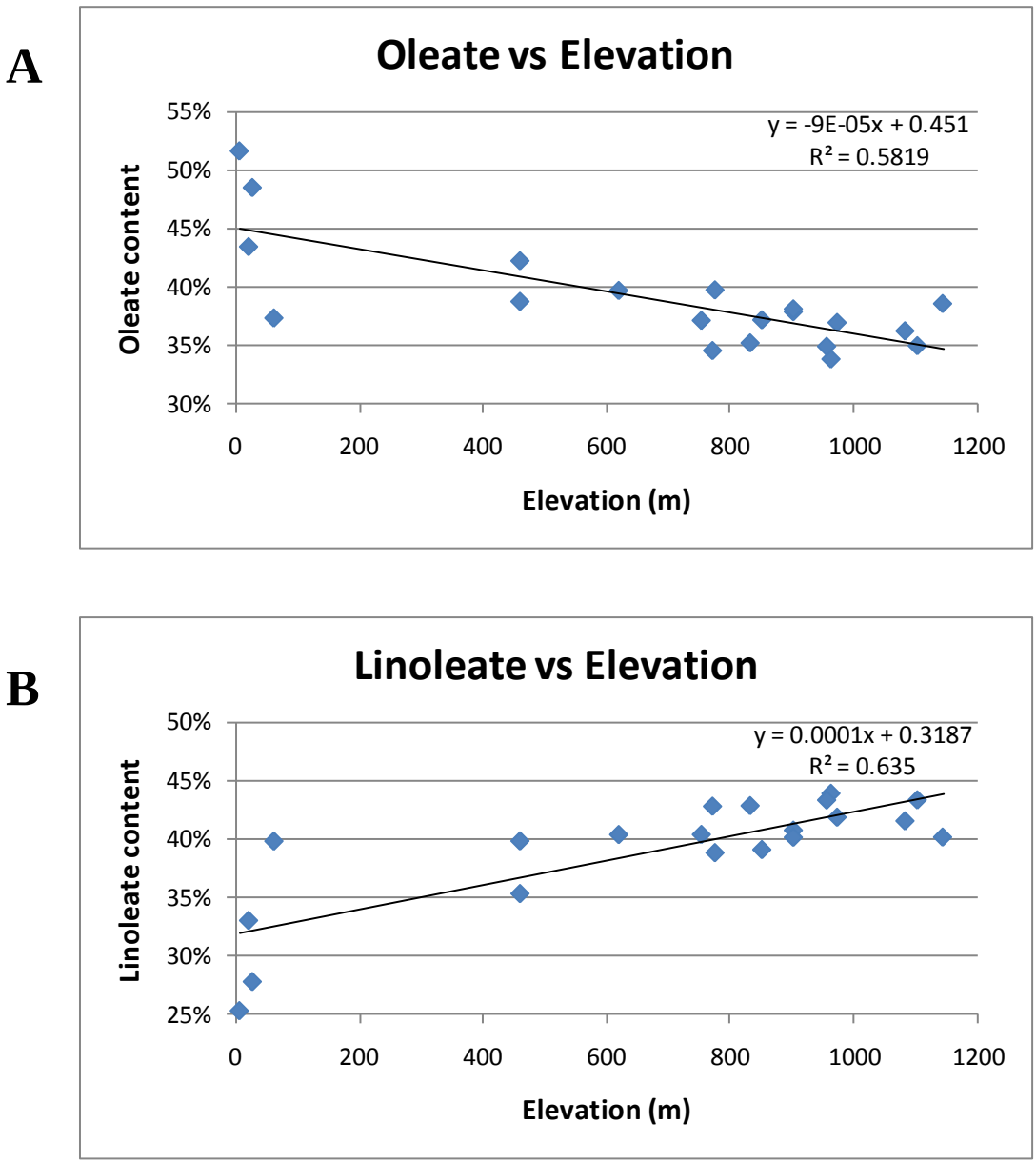


Figure 3.5 Plot comparison of oleate (A) and linoleate (B) composition against elevation (Table 3.1) in 23 Madagascan provenances.

3.2.3 Presence of phorbol esters is a discrete trait

To distinguish the edible and non-edible varieties of *J. curcas* seeds, PE content was measured in the seed samples in 37 provenances from Madagascar, Mexico, and two commercial suppliers from Tanzania and Suriname. HPLC determination of PE content revealed that all samples obtained from outside Mexico contain PEs within the range of 0.12 to 0.51 mg g^{-1} kernel weight (Figure 3.6). Most of the samples from Mexico lack PEs, with the only exception that the sample from the state of Chiapas contains 0.27 mg g^{-1} PEs (designated as sample ROS, Figure 3.6).

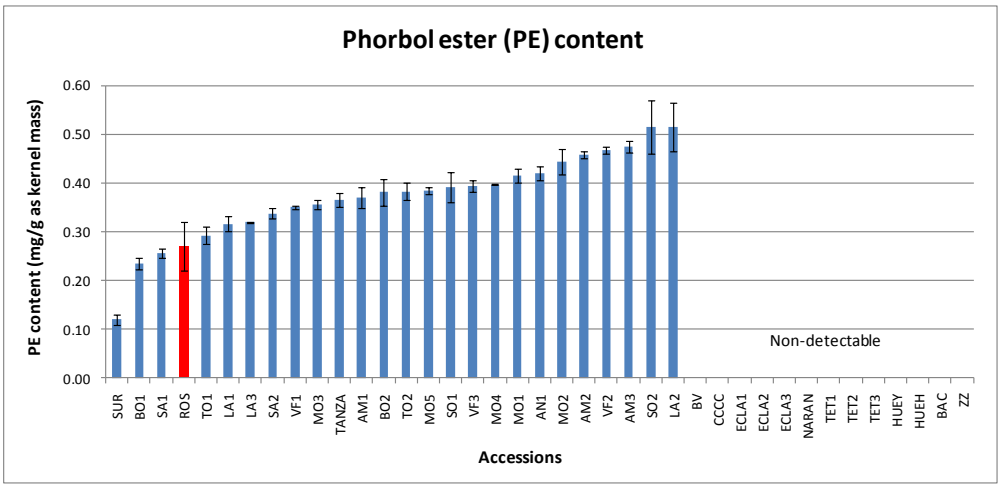


Figure 3.6 Phorbol ester (PE) content of *J. curcas* seeds collected from Madagascar, Mexico, and two commercial suppliers in Tanzania and Suriname. PE concentrations are calculated with reference to the internal standard phorbol 12-myristate-13-acetate (PMA). Error bars represent ± standard error mean of 3 technical replicates. Provenances codes are given in Table 2.1 in Chapter 2. Red bar indicates the only Mexican sample which contains PEs.

3.3 Characterization of oil mobilization during seed germination and seedling establishment

3.3.1 Water, oil content and fatty acid composition during germination and seedling establishment

During seed germination and seedling establishment, *J. curcas* seeds absorb a large amount of water to initiate a variety of biochemical activities and use lipids reserves to fuel this process. In order to characterize seed germination and seedling establishment, water and FA contents were measured during these biological processes in *J. curcas* seeds. 24 hours after imbibition, seeds were set up for germination and examined for a 6 days period. The rigid black testa ruptured at 1 DAI (day-after-imbibition), then the hypocotyl elongated and the radicle protruded at 2 DAI. This process is referred to as germination (see Chapter 2, Figure 2.1). From 2 DAI, the hypocotyl and radicle continued elongating and, at 6 DAI, seedlings were transferred to soil. At this stage the endosperm is attached to the cotyledon, providing the energy needed for further postgerminative growth using endospermic lipids. In order to study oil breakdown for the full duration of the germination process, samples in the wizened endosperm stage were also included. From 2 DAI to wizened endosperm stage, the process is referred to as seedling establishment. Water content changes greatly during imbibition, germination and seedling establishment (Figure 3.7A). Only approximately 7% of water is present in both embryo and endosperm in the dry seed. During imbibition, water content increases to 31% and 35% in the embryo and endosperm, and keeps increasing until 3 DAI. After 3 DAI water content in the seedling (79%) starts to decrease until the endosperm wizens (71%), whereas in the endosperm (59%) it increases gradually until the endosperm wizens (86%). When cotyledons spread the cotyledons contains 84% of water; this decreases again to 71% when the true leaf comes out. The water content in vegetative tissues is 86% in the true leaf, 92% in the stem, and 83% in the root (data not shown). The high water content in vegetative tissues in *J. curcas* is consistent with that of a succulent plant (Heller, 1996).

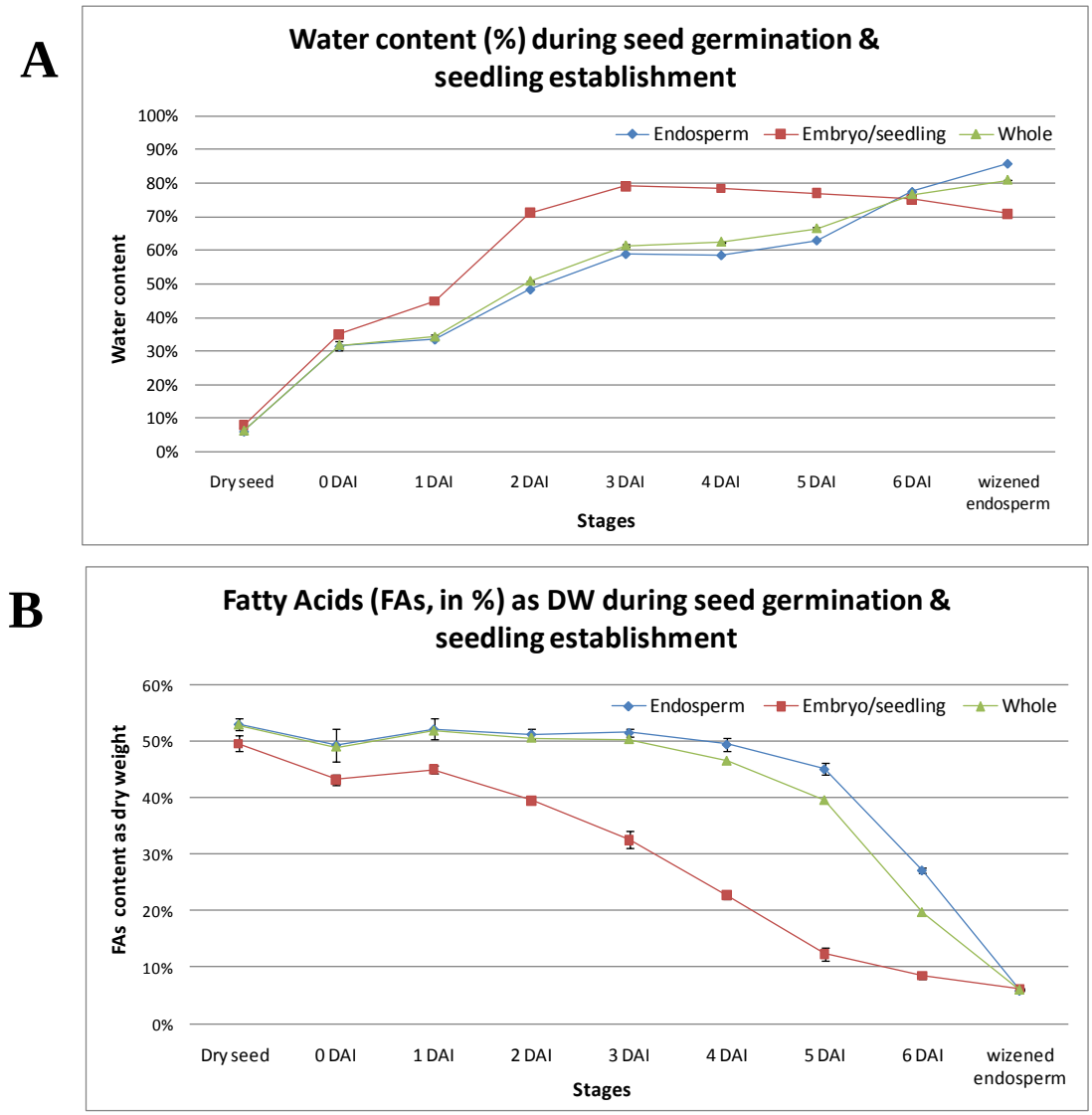


Figure 3.7 Water content (A) and oil content content (B) during seed germination and seedling establishment. DAI: days after imbibition; DW: dry weight; FAs: fatty acids. In panel B error bars show standard error of the mean of 3 technical replicates.

In *J. curcas* seeds about 95% of the oil is found in the endosperm with the remainder in the embryo. Thus the endosperm is the main provider of the energy needed for germination and seedling establishment. In terms of fatty acid content (percent as dry weight) the embryo contains slightly less than the endosperm at the dry-seed stage (Figure 3.7B). While the embryonic lipids are consumed earlier during germination from 1 DAI, the endospermic lipid levels only begin to fall after 3 DAI. At 5 DAI, only 12% oil is left in the seedling. FA content in the endosperm starts to diminish slightly between 3 DAI and 4 DAI, followed by a rapid decrease observed from 5 DAI. This indicates that *J. curcas* seeds use mainly embryonic lipids first for germination and initial post-germinative growth, and employ endospermic lipids at the later stages to support further seedling establishment.

The changes of FA composition in the endosperm and embryo were also analyzed (Table 3.2). From dry seed to 6 DAI, oleate and linoleate are the most abundant fatty acids in both endosperm (78% in dry seed; 77% in 6 DAI) and embryo (77% in dry seed; 63% in 6 DAI). During this period, in the endosperm the fatty acid composition does not change much, and only changes slightly when endosperm wizens. This implies that different FAs within the endosperm are metabolized at an equal rate throughout germination and seedling establishment. However, in the embryo there is a remarkable alteration in the fatty acid composition. From 3 DAI stearate, oleate and linoleate start to decline; linoleneate (18:3) begins to accumulate earlier at 2 DAI and constitutes the main fatty acid after cotyledons spread. This higher linoleneate composition is also observed in the true leaf and stem.

Table 3.2 Fatty acid composition in *J. curcas* seed during germination and seedling establishment, and in the vegetative tissues. E+S: endosperm + seedling.

Stages	Tissues	% fatty acid composition						
		16:0	16:1	18:0	18:1	18:2	18:3	20:0
Dry seed	Endosperm	14.7%	0.9%	5.6%	39.3%	39.1%	0.2%	0.2%
0 DAI	Endosperm	14.8%	1.0%	5.5%	39.8%	38.7%	0.2%	0.2%
1 DAI	Endosperm	14.5%	0.9%	5.7%	39.0%	39.5%	0.2%	0.2%
2 DAI	Endosperm	14.4%	0.9%	6.1%	41.0%	37.1%	0.2%	0.2%
3 DAI	Endosperm	14.6%	0.9%	5.9%	39.4%	38.6%	0.3%	0.2%
4 DAI	Endosperm	14.8%	0.9%	5.9%	38.5%	39.5%	0.3%	0.2%
5 DAI	Endosperm	15.7%	1.0%	5.9%	37.1%	39.6%	0.4%	0.2%
6 DAI	Endosperm	15.2%	1.0%	6.6%	39.5%	36.8%	0.6%	0.3%
Wizened endosperm	Endosperm	16.0%	0.7%	7.7%	38.6%	35.1%	1.4%	0.5%
Dry seed	Embryo	12.1%	0.3%	9.7%	34.2%	42.6%	0.8%	0.3%
0 DAI	Embryo	11.6%	0.3%	10.1%	33.4%	43.4%	0.9%	0.4%
1 DAI	Embryo	11.6%	0.2%	10.0%	34.1%	42.8%	0.8%	0.4%
2 DAI	Seedling	11.4%	0.3%	11.1%	35.1%	39.6%	2.2%	0.4%
3 DAI	Seedling	11.5%	0.3%	10.8%	31.6%	41.7%	3.7%	0.4%
4 DAI	Seedling	11.9%	0.3%	9.9%	30.0%	43.2%	4.3%	0.5%
5 DAI	Seedling	13.0%	0.3%	9.4%	26.5%	42.7%	7.6%	0.5%
6 DAI	Seedling	14.1%	0.4%	8.7%	25.0%	38.2%	13.1%	0.5%
Wizened endosperm	Seedling	14.5%	0.3%	7.9%	18.3%	34.7%	23.7%	0.6%
Dry seed	Whole seed	14.6%	0.9%	5.8%	39.0%	39.3%	0.2%	0.2%
0 DAI	Whole seed	14.6%	0.9%	5.8%	39.4%	39.0%	0.2%	0.2%
1 DAI	Whole seed	14.4%	0.9%	5.9%	38.7%	39.7%	0.2%	0.2%
2 DAI	E + S	14.2%	0.9%	6.4%	40.7%	37.3%	0.3%	0.2%
3 DAI	E + S	14.4%	0.9%	6.2%	38.9%	38.8%	0.5%	0.2%
4 DAI	E + S	14.4%	0.8%	6.3%	37.6%	39.9%	0.7%	0.2%
5 DAI	E + S	15.2%	0.9%	6.5%	35.3%	40.1%	1.6%	0.3%
6 DAI	E + S	14.7%	0.7%	7.4%	33.7%	37.3%	5.6%	0.4%
Wizened endosperm	E + S	15.2%	0.5%	7.8%	28.4%	34.9%	12.5%	0.5%
Cotyledon spread	Cotyledon	15.8%	0.3%	5.9%	7.3%	26.0%	44.1%	0.6%
True leaf	Cotyledon	15.4%	3.1%	4.3%	1.8%	13.7%	61.1%	0.5%
	True leaf	22.0%	1.1%	2.5%	3.4%	22.4%	48.2%	0.4%
	Stem	23.3%	0.2%	3.2%	3.8%	24.0%	44.7%	0.7%
	Root	23.9%	0.2%	3.9%	11.7%	35.4%	24.1%	0.8%

3.3.2 MLS and ICL are expressed during oil mobilization

To better characterize the oil mobilization in *J. curcas*, transcript abundance of two genes involved in oil mobilization in germination were analyzed. These genes encode two key enzymes of the glyoxylate cycle: malate synthase (MLS) and isocitrate lyase (ICL) (Graham, 2008). Using a 454 transcriptome database of developing seeds (King et al., 2010), the *J. curcas MLS* and *ICL* genes were identified and then amplified (data not shown), and their transcript levels were analyzed by qRT-PCR from dry seed to wizened endosperm stages. *MLS* transcript accumulation in endosperm decreases transiently after imbibition, peaks again at 3 DAI and is expressed stably to the wizened endosperm stage with gradual diminishing (Figure 3.8A). In embryo/seedling, *MLS* expression is relatively lower than in the endosperm, and decreases gradually from dry seed to the wizened endosperm stage. With regard to *ICL*, transcript accumulation in endosperm increases at 2 DAI and remains high until the wizened endosperm stage, peaking at 5 DAI (Figure 3.8B). In embryo/seedling *ICL* is expressed at lower levels than in endosperm and, contrary to the pattern in endosperm, its expression decreases further from 2 DAI.

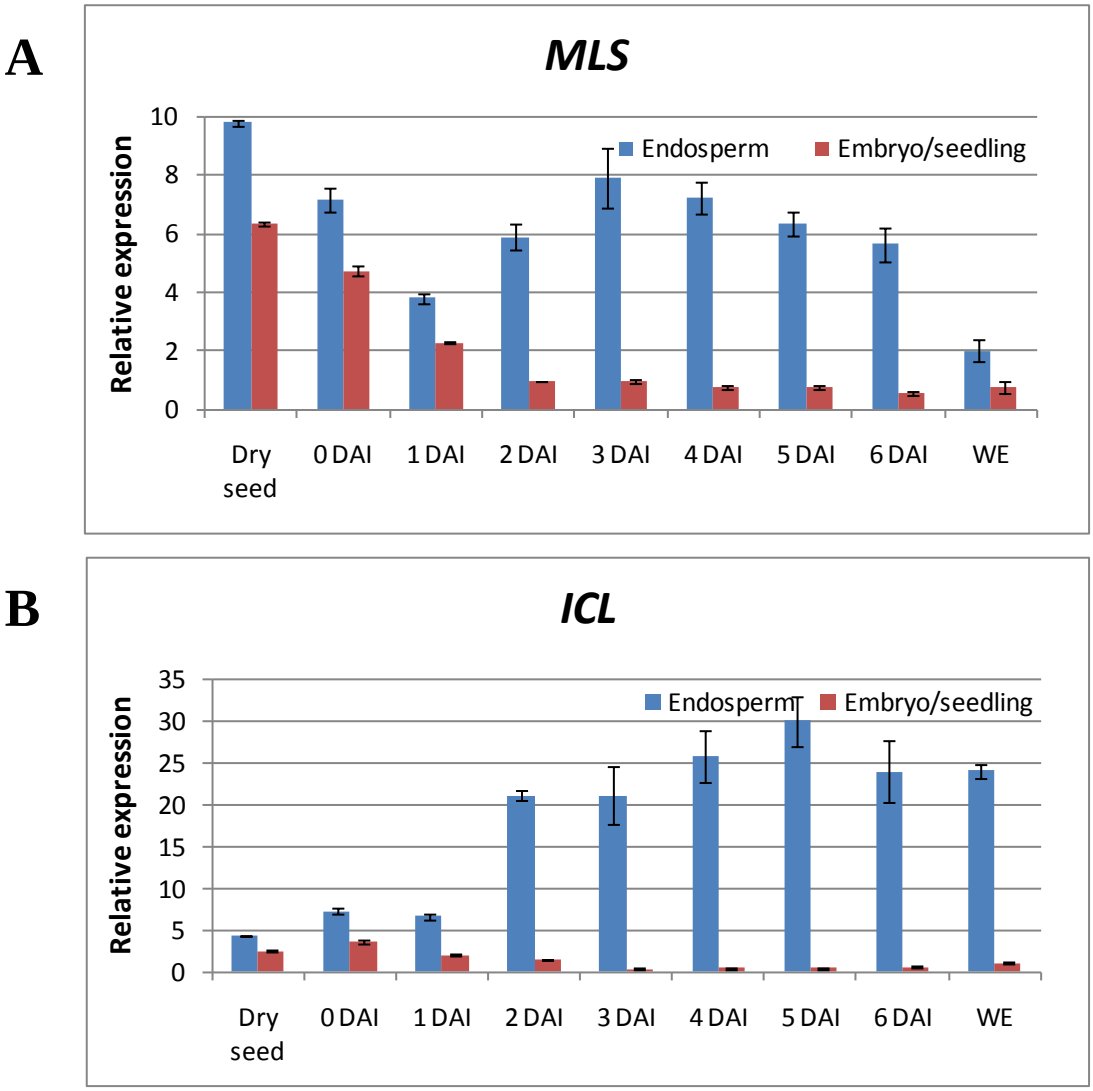


Figure 3.8 Gene expression of *MALATE SYNTHASE* (*MLS*) and *ISOCITRATE LYASE* (ICL) in *J. curcas* seed during germination and seedling establishment. Gene expression levels are normalized to *ACTIN*. WE: wizened endosperm. DAI: days after imbibition. Error bars represent standard error of the mean of 3 technical replicates.

3.4 Discussion

3.4.1 Oil content and fatty acid composition of J. curcas seeds

The oil content of *J. curcas* seeds of 23 provenances from Madagascar was analysed in this study. The average oil content was between 24.0% and 37.6%, and also the high variation in oil content is in accordance with previous findings (Jongschaap et al., 2007). The oil content of *J. curcas* is of vital importance for the evaluation of its economic value. Therefore an accurate measurement of the seed oil content is

necessary for any claims of high oil content variety, from either the selection of naturally occurring accessions or the development of new bred lines.

From this study, no correlation was found between the seed mass, kernel mass, kernel percentage and the oil content in *J. curcas* seed in the material analyzed. In addition, neither altitude nor the soil composition was observed to have a direct impact on the seed oil content. One possible explanation for this is that the oil content data from only 23 provenances were available in the present analysis; this is far too few when as many as 18 variables were used in the PCA analysis and makes it difficult (if not impossible) to detect any correlation. In addition, it is almost certain that the oil content depends on one or more environmental factors, which were not available in the present analysis. The age of *J. curcas* plant and the site average annual rainfall affect the oil content (Achten et al., 2008).

In terms of oil composition, the two unsaturated FAs oleate and linoleate account for around 75% of the total fatty acids in all the samples analyzed. It is interesting to note that the ratio of these two FAs varied considerably, and that the concentrations of these FAs correlated significantly with the altitude. It has been shown in other oilseed crops-such as soybean and sunflower seeds-that temperature has an effect on the FA composition with desaturation increasing as temperature decreases (Wolf et al., 1982, Harris et al., 1978). Given the assumption that the mean annual temperature decreases with elevation, it can be hypothesized that temperature plays an important role in FA composition in *J. curcas* seed oil as well.

Biodiesel is currently defined as monoalkyl-esters of long-chain fatty acids from plant or vegetable oils (Tyson, 2009). The EU and USA have their own standards of specification which biodiesel must meet, which are EN14214 and ATSM D6751, respectively (Table 3.3). Many of the key properties of biodiesel, including the cetane number (CN), cold-flow, viscosity and oxidation stability are determined by the FA composition of the feedstock vegetable oil (Table 3.3) (Knothe, 2008, Tyson, 2009).

Table 3.3 Specifications for key biodiesel properties and how these properties are affected by the FAs of the feedstock. Min: minimum. 1: not specified; 2: need to report to customers; 3: at 40 °C ($mm^2 s^{-1}$); 4: hour at 110 °C. Table adapted from Knothe (2008) and Tyson (2009).

Specification for biodiesel			Biodiesel properties related to the FAs of the feedstock		
	EN 14214	ASTM D6751	Saturated	Mono-unsaturated	Poly-unsaturated
Cetane number (CN)	51 min	47 min	high	medium	low
Cloud point	N/A[1]	report[2]	high	medium	low
Kinematic viscosity[3]	3.5–5.0	1.9–6.0	high	medium	low
Oxidative stability[4]	6 min	3 min	high	medium	low

CN is perhaps the most important factor among these properties. CN increases with an increasing monoalkyl-esters chain length and decreases with unsaturation of FAs (Knothe, 2008). EN 14214 and ASTM D6751 requires a minimum CN of 51 and 47, respectively. Above a CN value of 54.5, there are no improvements in engine performance (Icngur and Altiparmak, 2003). In this study all of the oil samples from Madagascar had a fatty acid composition which produced a calculated CN that would meet the stricter EU requirement, and thus be suitable for producing biodiesel. However, it is noticeable that many of the linoleate-rich samples gave a calculated CN of 51, which is close to the threshold value for EN 14214. The calculated CNs in this study are in agreement with actual measured CN of *J. curcas* biodiesel, which have been found to fall within the range of 50-57 [reviewed by (Achten et al., 2008) and (King et al., 2009)]. In summary, in most instances, *J. curcas* biodiesel is likely to have a CN value which meets both American and European standards. Where CN values are too low, the CN value can be increased by blending, through use of ethyl esters rather than methyl-esters, or by adding cetane enhancers (Knothe, 2008, Foidl et al., 1996).

Cold-flow of biodiesel is determined largely by the concentration of saturated FA composition (Imahara et al., 2006). The saturated FA content of *J. curcas* oil is typically 20% [this study and Foidl et al (1996)]. The low-temperature operability of biodiesel is most commonly measured by the cloud point, at which temperature small solid crystals start to form (Tyson, 2009). The cloud point of *J. curcas* oil FAMEs has been determined as 4–8 °C (Krishnakumar et al., 2008, Sarin et al., 2007). This temperature range is too high for using 100% biodiesel (B100) in temperate climates during winter, as the saturated FAMEs would be prone to

crystallization at lower temperatures (Srivastava and Prasad, 2000). However, at present most biodiesel is typically sold as blends of petrodiesel (e.g., B5 which contains 5% biodiesel). The cloud-point of these blends is much lower than B100. Cold-flow properties can also be improved by a winterization process, which removes the high-melting point saturated esters (Dunn et al., 1996, Lee et al., 1996). Alternatively, transesterification of fatty acids with branched chain alcohols (such as isopropanol) instead of methanol has also been shown to produce biodiesel with a lower cloud-point (Lee et al., 1995).

The viscosity of biodiesel increases with the length of monoalkyl-ester and decreases with unsaturation (Knothe, 2008). The kinetic viscosity of biodiesel specified by EN 14214 and ASTM D6751 is 3.5-5.0 and 1.9-6.0 $mm^2 s^{-1}$ at 40° C, respectively. The kinetic viscosity of FAMEs produced from *J. curcas* oil of different FA composition lies within this range (Achten et al., 2008, Anwar et al., 2010).

Oxidative stability of biodiesel is greatly affected by the unsaturated FAs, in particular, the polyunsaturated FAs. Double bonds in the unsaturated FAs are susceptible to reaction with oxygen. No previous study has reported the oxidative stability of *J. curcas* biodiesel, though the iodine number, which reflects the unsaturation of *J. curcas* FAMEs, falls within the range of the requirements of EN 14214 and ASTM D6751 (Achten et al., 2008). However, the study by Knothe (2008) showed that the most oxidatively stable unsaturated fatty ester-methyl oleate has an oxidative stability of 2.68, which fails to meet either the EN 14214 (min 6) or the ASTM D6751 (min 3) specification. Thus to meet these specifications, an antioxidant is necessary for any biodiesel containing unsaturated esters. The use of antioxidant additives is common to improve the oxidative stability of biodiesel. Other properties of *J. curcas* oil FAMEs as biodiesel have been reviewed by Achten et al (2008), King et al (2009) and Brittaine and Lutaladio (2010), and have shown satisfactory results to qualify for the EU and USA criteria.

In conclusion, *J. curcas* oil is suitable to be transesterified into biodiesel. Improvement in seed oil content will increase the economic value of *J. curcas*. Some processing would be necessary to improve the cold-flow and oxidative stability. Genetic engineering could be a useful tool to improve oil quality and

quantity. *Acetyl-CoA carboxylase* (*ACC*), *diacylglycerol acyltransferase* (*DGAT*) could be the possible target to increase oil content (Buchanan et al., 2000). Since mono-unsaturated fatty acids in the feedstock oil are desirable for biodiesel production (in *J. curcas* mainly oleate), the oleate content can be increased by disrupting genes such as the *fatty acid desaturase2* (*FAD2).* Genetic engineering has already been successfully used to modify fatty acid content in seed oil including for example in rapeseed to reduce erucic acid (Buchanan et al., 2000).

3.4.2 Phorbol esters content

To date *J. curcas* plants producing PE-free seeds have been found only in the Mexican states of Vera Cruz (Makkar et al., 1998b), Quintana Roo (Makkar et al., 1998a) and Puebla (King et al., 2009). Seeds from these areas are regularly consumed by the indigenous people after roasting (Schmook and Serralta-Peraza, 1997, Gubitz et al., 1999). Animal feeding experiments using edible seed meal also did not show adverse effects (Makkar et al., 1998a, Makkar et al., 1998b). It is generally believed that seeds outside Mexico contain PEs (Brittaine and Lutaladio, 2010), and also in Mexico, PE containing seeds have been reported from the states of Vera Cruz and Morelos (Makkar et al., 1998a). Within the collection in the present study, PEs were found in all the seed samples sourced outside Mexico, and from the Mexican state of Chiapas. The presence of PEs in the sample from Chiapas confirms that non-edible seeds also occur in Mexico. PEs were not detected in seed samples obtained from the states of Puebla, Quintana Roo, and Vera Cruz in Mexico, which agrees with previous observations (Makkar et al., 1998a, Makkar et al., 1998b, Schmook and Serralta-Peraza, 1997). As *J. curcas* seeds are often consumed in Mexico, rigorous containment measures should be adopted in the cultivation of non-edible varieties to prevent accidental human consumption. The complete absence of detectable PEs in the edible variety suggests that seed edibility may be a qualitative trait. This would simplify introgression of this trait into other germplasm through plant breeding.

It is noticeable that the PE content in this analysis is approximately 5 to 10-fold lower compared to values reported by previous studies (Makkar et al., 1998a, Makkar et al., 1998b, Xiao et al., 2011). In both Makkar' previous studies and this

study, phorbol 12-myristate-13-acetate (PMA) was used as the internal standard to measure the phorbol ester concentration. However, in this study the peak area of PEs was measured at 240 nm rather than at 280 nm (used by Makker). The UV spectrum of PMA has a single peak with a λ_{max} of 242 nm, whereas the *J. curcas* PEs have mass spectra with multiple peaks (see Appendix 2). This includes a peak at 280 nm, and a peak in the region of 238-242 nm. Based on a comparison of these spectra, the peak in the 238-242 nm region is likely to be the phorbol region of the spectra, and the other peaks appear as a consequence of the ring-structures present in the intramolecular diesters (Haas et al., 2002), which PMA lacks (Figure 3.2, PMA and PE 1-6). Previous calculations using peak areas at 280 nm of PMA are therefore likely to have significantly overestimated the concentration of *J. curcas* PEs.

Since no legal limit has been established regarding the acceptable level of PEs, as the main toxic component in *J. curcas* seed meal, PEs should be completely removed from any seed meal to be used as animal feed. This could be achieved by physiochemical means (Section 1.4.4) or by down-regulating PEs biosynthesis, or by taking advantage of naturally occurring edible varieties. The PEs biosynthesis pathway in *J. curcas* is still not fully understood, but progress has been made in our labotory (Dr. Andy King, unpublished results). Once the key genes in this pathway have been characterized, it would be possible to develop new *J. curcas* varieties without PEs in the seed. However, development of a marker linked to phorbol-ester production will allow breeding of new edible varieties.

3.4.3 Lipid breakdown during seed germination and seedling establishment

In the present study, tissue specific profiles of lipid breakdown were observed during seed germination and seedling establishment. Oil mobilization occurs in the embryo prior to the endosperm after the imbibition. This pattern in *J. curcas* contrasts with the situation in *Arabidopsis*, where the endospermic lipid is consumed first during germination and embryonic lipid is utilized for later stages (Penfield et al., 2006). This difference correlates with the fact that in *Arabidopsis* lipids are stored primarily in the embryo whereas in *J. curcas* the endosperm constitutes the majority of the seed. In castor, the lipid content remains unchanged in the endosperm during early

periods of germination (Muto and Beevers, 1974), and this pattern is also found in the endosperm tissue of *J. curcas* (Yang et al., 2009).

During the seedling establishment, at 5 DAI, lipid in the embryo has almost depleted (12% FAs), and the endospermic lipid starts to break down at a considerable rate to provide the seedling with supplies of carbohydrate for the post-germinative hypocotyl growth. Although at 2 DAI germination has completed in *J. curcas* (with the radicle emergence), the endosperm is still essential for the post-germinative seedling establishment, since the seedling is still not capable of photoautotrophism. In *Arabidopsis*, the endosperm only stores about 10% of the total lipid in the seed, but if it is removed hypocotyl elongation is severely reduced in the dark (Penfield et al., 2004). The glyoxylate cycle and glyconeogenesis convert the lipid in *J. curcas* seed endosperm, and carbohydrate must be transported to the seedling to exert this function.

During germination and seedling establishment in *J. curcas* the endosperm uses all the FAs non-selectively, with only a very minor decrease in the linoleate content in later post-germinative stages. This observation is in agreement with the previous study by Yang et al (2009), but it is noticeable that in their study a more significant decrease in the linoleate content was observed. In castor, which is similar to *J. curcas* in that the endosperm tissue stores most of the oil reserve, the oil mobilization shows no preference in using the FAs. A slight decrease in the ricinoleic acid content occurs only when it is 6 days after germination (Marriott and Northcote, 1975). In the embryo tissue in *J. curcas* oleate (and linoleate in the later stages) content decreases remarkably, indicating that lipid mobilization selects these two FAs preferentially. This compares with the corresponding increase of linoleneate (and to a lesser extent palmitate), which most probably reflects the *de novo* synthesis and accumulation of these FAs in galactolipids as the seedling becomes photosynthetically competent.

The expression patterns of two genes encoding key enzymes in the glyoxylate cycle, *MLS* and *ICL* were studied. In the embryo, after imbibition, the expression levels of both *MLS* and *ICL* decrease from a moderate level and keep at a basal level after 3 DAI. This relatively high expression of *MLS* and *ICL* before 3 DAI might explain

the rapid lipid breakdown in the embryo in *J. curcas* from 1 DAI. In the endosperm, although expressed in different patterns from dry seed to 1 DAI, both *MLS* and *ICL* are expressed at a high level from 2 DAI, which reflects the lipid breakdown in the endosperm starting from 3 DAI. The coordinate gene expression and enzyme activity for *ICL* and *MSL* genes have been demonstrated in germinating cotton (Turley and Trelease, 1990) , cucumber (Weir et al., 1980) and *Arabidopsis* (Rylott et al., 2001). It is interesting that the expression of *MLS* and *ICL* in *J. curcas* seeds exhibit different patterns. In *J. curcas* Yang et al (2009) analyzed the temporal protein profiles in the endosperm during germination. Proteins with a 2-fold change (up- or down-regulated) in the endosperm from 0 h to 96 h after imbibition (which is likely equivalent to 5 DAI in this study) were analyzed. A steady increase up to 6-fold of ICL during germination was observed, which is in agreement with the change of the transcript abundance of *ICL* in the present study (about 6-fold change as well). However, MLS was not analyzed in Yang's protein profiling study (presumably due to the change being less than 2-fold). This might corroborate the expression pattern of *MLS* revealed in the present study—from 0 DAI to 5 DAI the transcript abundance of *MLS* changes less than 2-fold. To conclude, a distinct lipid breakdown pattern in the endosperm and embryo in *J. curcas* has been observed, which is in line with the differential expression of two key genes in the glyoxylate cycle, *MLS* and *ICL*.

Chapter 4: Genetic profiling of *J. curcas* using AFLP markers

4.1 Introduction

4.1.1 Why AFLP markers and the rationale for their use

Several different molecular marker based strategies can be used for genotyping, including RAPD (Random Amplification of Polymorphic DNA), ISSR (Inter-Simple Sequence Repeat), AFLP (Amplified Fragment Length Polymorphism) , SSR (Simple Sequence Repeat) and SNP (Single-Nucleotide Polymorphism) (Weising et al., 2005). A comparison of those molecular marker techniques, concerning their reproducibility, dominant/codominant nature and other properties is shown in Table 4.1. Among these marker systems AFLP is especially useful for species where *a priori* DNA sequence information is not available, and its high reproducibility (>98%) has been well proven (Jones et al., 1997, Myburg et al., 2001). Nowadays next generation DNA sequencing such as 454 pyrosequencing (Rothberg and Leamon, 2008) and Illumina sequencing (Morozova and Marra, 2008) have significantly accelerated the process of acquiring ESTs or genome sequences from species of interest, and these sequencing techniques have been applied on *J. curcas* research to obtain extensive DNA sequence information or to assist marker development (Sato et al., 2011, Yadav et al., 2010). However prior to 2007, there were very few sequences of *J. curcas* deposited in the public domain, and likewise there were few relevant genetic diversity studies. At that time only two studies employing RAPD and ISSR markers and reporting on the genetic diversity in *J. curcas* had been published (Sujatha et al., 2005, Basha and Sujatha, 2007). Therefore the AFLP marker approach was selected at the outset of the current project to assess the genetic diversity in a collection of *J. curcas* germplasm available to the Graham laboratory at the University of York.

As a second generation DNA fingerprint methodology developed in 1995 (Vos et al.), the AFLP marker technique has broad applicability in many studies in various organisms including plants, animals and bacteria [reviewed by Meudt and Clarke (2007), Bensch and Akesson (2005)]. It is a versatile tool and particularly in plants, it has been widely used in genotyping and genetic diversity studies (Krauss, 2000,

Lanteri et al., 2004, McLenachan et al., 2000), marker assisted selection (MAS) for improved yields, traits, or disease resistance (Ajmone-Marsan et al., 2001, Baszczyk et al., 2005, Schierholt et al., 2000, Yin et al., 1999), and also for the construction of genetic maps (Alonso-Blanco et al., 1998, Ipek et al., 2005, Lionneton et al., 2002). Compared to other marker systems, apart from the advantages mentioned above, AFLP provides a considerable number of multi-loci markers. Though AFLP analysis is less informative at a given locus due to its dominant nature, it is still a method of choice since other molecular markers are still limited in number.

Table 4.1 A comparison of five different molecular marker techniques. Summarized from "DNA fingerprinting in plants: principles, methods, and applications" (Weising et al., 2005).

Characteristics	RAPD	ISSR	AFLP	SSR	SNP
Prior sequence information required	N	N	N	Y	Y
Number of loci examined/primer	About 10	About 10	20-100	1	1
Inheritance pattern	Dominant	Dominant	Dominant	Codominant	Codominant
Level of polymorphism	Medium	Medium	High	High	High
Reproducibility	Low	Medium	High	High	High
Commercial package available	Y	Y	Y	N	N
Level of training required	Low	Low	Medium	Low/High[1]	High
Cost	Low	Low	Medium/ High[2]	Low/High[3]	High

1: If specific primers have to be developed.
2: If automated equipment needs to be purchased.
3: If specific primers have to be developed, and/or automated equipment needs to be purchased.

The experimental procedure involved in AFLP analysis is illustrated in Figure 4.1. In Step 1, genomic DNA is extracted and then double digested by restriction enzymes, including a frequent cutter (recognition site is 4-base pair, such as MseI) and a rare cutter (recognition site is 6-base pair, such as EcoRI). Complementary adaptors are ligated to the sticky ends of the DNA fragments in Step 2. Two successive PCR reactions are then performed, which are pre-amplification (Step3)

and selective-amplification (Step4). During the selective-amplification labelled EcoRI primers and unlabeled MseI primers bind to the matched fragments and generate 20-100 amplicons, which can be examined by electrophoresis. Polymorphisms occur when amplicons are present ("1") in some genotypes but absent in others ("0"). In brief, substitutions of nucleotides in the restriction sites or the selective primer binding sites, indels or variation of microsatellite between the restriction sites all contribute to the polymorphisms. Single nucleotide polymorphisms are likely to be the most frequent type of mutation causing the loss of a restriction site. The binary ("1" and "0") matrix can be analyzed by software to calculate polymorphisms, construct a phylogenetic tree and infer genetic structure. A comprehensive review on AFLP methodology and applications was performed by Meudt and Clarke (2007).

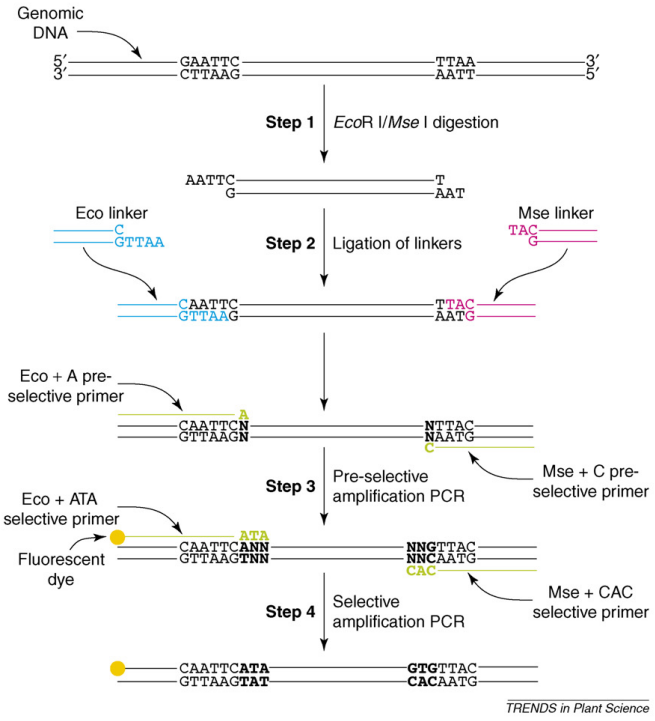


Figure 4.1 The schematic of AFLP marker technique. Figure taken from Meudt and Clarke (2007)

4.1.2 Factors considered for experimental design and data analysis

4.1.2.1 Marker development

Acquisition of the marker profile is the first and paramount step in AFLP analysis. This information (presence or absence of bands, scored as "1" or "0" matrix) serves as the original data for any downstream analysis, usually assisted by computer programs. In the experimental step, both positive and negative controls should be included to monitor and detect contamination and ensure reproducibility. Samples should be randomized to avoid any investigator-associated biases in gel scoring. Consistency of scoring is essential which is assisted by discarding all equivocal scores or recording as missing data. Details of the routines established for meticulous scoring are provided in Section 2.3.

To date most AFLP studies in plants have used a combination of EcoRI and MseI restriction enzymes (REs) possibly due to their well established usage, routine protocol, and relatively low cost (Bensch and Akesson, 2005). Few studies have focused on exploring the REs selection, which normally involves the optimization of the frequent cutter restriction enzyme. Due to the variation in the restriction sites, different REs behave differently in their ability to detect polymorphisms and the distribution of amplified fragments that they generate. A two-fold increase of detected polymorphisms have been reported when using TaqI as the frequent cutting RE by Ajmone-Marsan in cattle (1997). In this analysis both MseI and TaqI were used as the frequent cutting enzymes. Based on their restriction sites, it was expected that TaqI (5'-T↓CGA-3') would identify more polymorphisms than MseI (5'-T↓TAA-3'), and in addition generate bands with better distribution of amplicons on the gel.

4.1.2.2 Constructing a phylogenetic tree

Once the gel score is "translated" into a binary matrix, downstream analysis can be performed to infer genetic diversity and structure which is most commonly presented as a phylogenetic tree.

The Distance method and Character-based method are the two primary strategies for tree construction. The Distance method is regarded as fast and reliable, providing that statistical support, such as bootstrap values are available. Firstly the binary matrix comprising "1" and "0" from the gel score is computed into an allele frequency matrix, which is then converted into a pairwise distance matrix. Secondly this pairwise distance matrix (or a corrected matrix calculated based on this distance matrix) is created and used for calculating branching orders and lengths to construct the phylogenetic tree.

The commonly used distance methods include the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) and Neighbour Joining (NJ). A major drawback of UPGMA is that it assumes ultrametric branching, which means that all taxa have equal distance to the root. This assumption is possibly false because it is unlikely that all the pedigrees have the same rate of evolution from a common ancestor. Thus, NJ as described by Hall (2004) and Weising et al (2005) is considered a more appropriate algorithm for the construction of phylogenetic trees.

A common pitfall in drawing phylogenetic trees is that the trees are often rooted incorrectly. The root in a phylogenetic tree indicates from which point all the taxa have evolved, and thus shows the direction of evolution. Unless the researcher is certain about where to place the root, an unbiased radiation tree, should be provided, even though it might be less easy to understand/interpret (Hall, 2004).

"*One of the most important things to understand about the phylogenetic trees that we estimate is that they are almost certainly wrong*" (Hall, 2004). Evolution is such a complex process that it is virtually impossible for a tree to reflect it precisely. Best estimates for tree building require statistical analysis that allows confidence levels to be ascribed. With regard to the distance method, bootstrapping is often the method of choice to ascribe confidence values for nodes in phylogenetic trees (Efron et al., 1996). A 70% bootstrap cut-off value is considered as a strong indicator for a group (Efron et al., 1996, Felsenstein and Kishino, 1993).

In the current analysis, a NJ phylogenetic tree was built and presented both with and without root to explicitly explain the genetic structure in *J. curcas*. Bootstrap values were indicated to evaluate the robustness of the tree.

4.1.2.3 Bayesian inference

The Character-based methods include Parsimony, Maximum Likelihood, and Bayesian analysis. Principles and algorithms of these methods are described by Hall (2004) and Holder & Lewis (2003). In principle, these methods search for and evaluate the tree(s) which best fit the available data. These methods generate and test a large number of trees (for 100 taxa there can be $2*10^{182}$ possible unrooted trees!) which takes a lot of time. Among these methods, Bayesian inference is a Markov Chain Monte Carlo (MCMC) based approach for phylogenetic inference (Mau et al., 1999, Rannala and Yang, 1996). Bayesian inference which involves simultaneous phylogeny inference and probability measurement, is a relatively new approach that has been increasingly employed for phylogenetic studies in the last ten years, thanks to the availability of powerful desktop computers and servers (Hall, 2004, Holder and Lewis, 2003). In the current analysis apart from the NJ tree building mentioned above, Baysien analysis was conducted as a second approach to depict the genetic structure in *J. curcas*.

4.1.3 Aims

The aims of this part of the study are to:

1. Investigate the genetic diversity and structure of *J. curcas* germplasm available from various regions (Table 2.1, Chapter 2).
2. Establish if it is possible to distinguish between edible and non-edible varieties at the genetic level.
3. Evaluate the effectiveness of MseI and Taq I as frequent cutter restriction enzymes in the AFLP analysis in *J. curcas*.

Achievement of these aims would be useful for subsequent work, such as the preservation of core varieties for germplasm banks (Gepts, 2006), and the selection of parental lines for the production of mapping populations (Schneider, 2005). It is to be expected that the edible and non-edible varieties would be found to be

genetically different, and that more genetic diversity would exist in the Mexican region.

4.2 Development of AFLP markers

AFLP analysis was performed on DNA from plants grown from the 42 seed batches detailed in Table 2.1 in Chapter 2. For each seed batch, the analysis was conducted on up to 3-4 individual plants (a total number of 126 DNA samples) to check whether any diversity exists within the seed batch. Figure 4.2 shows a representative polyacrylamide gel electropherogram obtained using the primer combination E_ACA and T_CG (Table 2.2, Chapter 2). The arrows on the gel indicate some of the polymorphic amplicons that were observed during the gel scoring. For the AFLP analysis, twelve primer combinations were used and yielded 337 unambiguous bands between 50 bp and 700 bp.

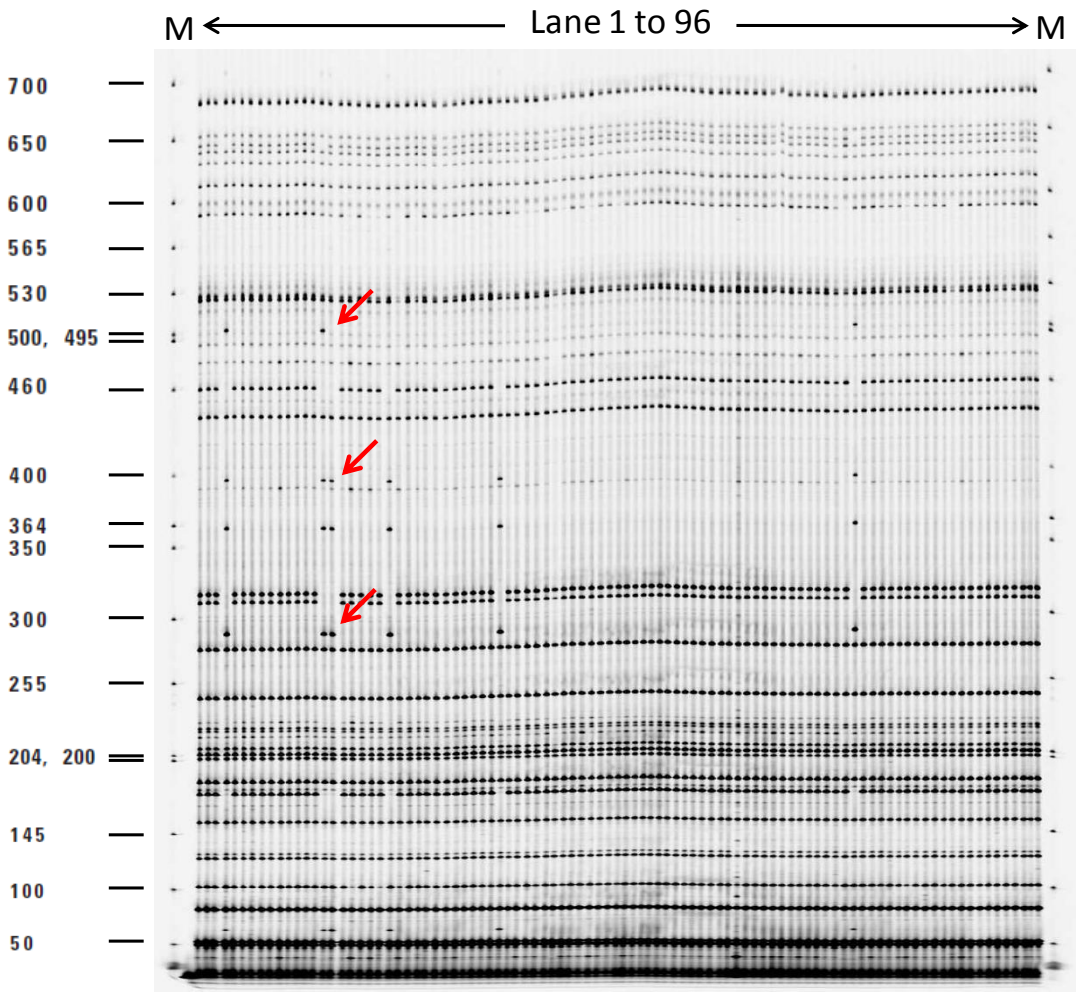


Figure 4.2 AFLP polyacrylamide gel using E_ACA+T_CG primers. Ninety six samples were randomized and analyzed, including 1 no template control (NTC) and replicates of 8 different samples as quality controls for reproducibility. PCR products of selective amplification were denatured and loaded (1.2 µl) onto the gel, alongside 50-700 bp DNA markers (1.0 µl). The red arrows on the gel indicate representative polymorphic markers.

4.3 Genetic diversity in *J. curcas* is high in Mexico but limited in other provenances

Out of a total number of 337 bands, 87 were polymorphic [PPB (percentage of the polymorphic bands) =25.8%]. In order to elucidate the genetic diversity in *J. curcas*, different samples were grouped in light of either their edibility or geographical origins, and within-group genetic diversity was calculated (Table 4.2).

In the upper panel of Table 4.2, within the non-edible seed samples (see Table 3.6 in Chapter 3 for the seed edibility), the PPB is 20.5%. This can be largely attributed to the Rosario Izapa samples from Mexico, which yielded 58 polymorphic bands (PPB=17.2%). When the samples from Rozario Izapa are excluded, genetic variation within the rest of the non-edible samples is extremely low (PPB=3.3% overall and PPB=1.5% within Madagascar). For a large subset of these samples, no polymorphisms were observed, suggesting that these materials are almost clonal. Within the edible Mexican samples, 30 polymorphic bands were detected (PPB=8.9%).

In the lower panel of Table 4.2, according to the geographical origins, comparison was made between the Mexican samples (including all edible samples and non-edible Rosario Izapa) and the non-Mexican samples. The results suggested that there is very little genetic diversity in the materials obtained from outside Mexico, for which only eleven bands are polymorphic (PPB=3.3 %). Variation is even lower in Madagascar alone—only 5 polymorphic bands were observed (PPB=1.5%). Within the Mexican samples genetic diversity is much higher with a total of 85 polymorphic bands being observed (PPB 25.2%). However as mentioned above, again this can be traced to the samples from Rosario Izapa (PPB=17.2%). Nevertheless, genetic variation in the edible Mexican samples is still comparatively higher (PPB=8.9%) than the non-edible samples outside Mexico (PPB=3.3%).

Table 4.2 Number of polymorphic bands observed in different *J. curcas* samples. PPB: percentage of the polymorphic bands.

Sample origin	Polymorphic Bands (x/337)	PPB
By seed edibility		
Non-edible samples	69	20.5%
Rosario Izapa	58	17.2%
Non-Mexican	11	3.3%
Madagascar	5	1.5%
Edible samples	30	8.9%
By geographical origin		
Mexican samples	85	25.2%
Edible Mexican	30	8.9%
Rosario Izapa	58	17.2%
Non-Mexican samples	11	3.3%
Madagascar	5	1.5%

4.4 Edible *J. curcas* provenances are genetically distinct from non-edible provenances

4.4.1 Genetic structure analysis using distance methods

In this study, the NJ method was employed to build the tree for analyzing AFLP markers. Figure 4.3 shows the midpoint rooted NJ phylogenetic tree constructed from a Nei's distance matrix (Nei, 1973). Bootstrap values (1000 replicates) were calculated to evaluate the robustness of the phylogenetic tree. According to the tree, all the samples separate into two clusters. All the non-edible samples from Madagascar, Brazil, India, Ghana, Tanzania, Suriname and Mexican Rosario Izapa, are contained within one cluster and all the edible samples from Mexico are contained in a second cluster.

In the non-edible clade, apart from the Mexican Rosario Izapa (R.I, shown in red), the entire non-edible samples group together (shown in blue; bootstrap value=100). It is worth mentioning that though R.I samples groups into this non-edible cluster, they exhibit strong divergence according to the branch length. This is also supported by the greater genetic variation shown in Section 4.3 and the genetic distance from R.I samples to other non-edible samples.

In order to strengthen the observation mentioned above, an unrooted radiation NJ tree without sample names is shown (Figure 4.4). The unrooted tree in Figure 4.4 shows that the edible samples from Mexico (shown in yellow) and the non-edible samples from outside Mexico (shown in blue) cluster into their own groups with very low within-group genetic variation, whereas Rosario Izapa (shown in red) is distinct from these two groups with high within-group variation.

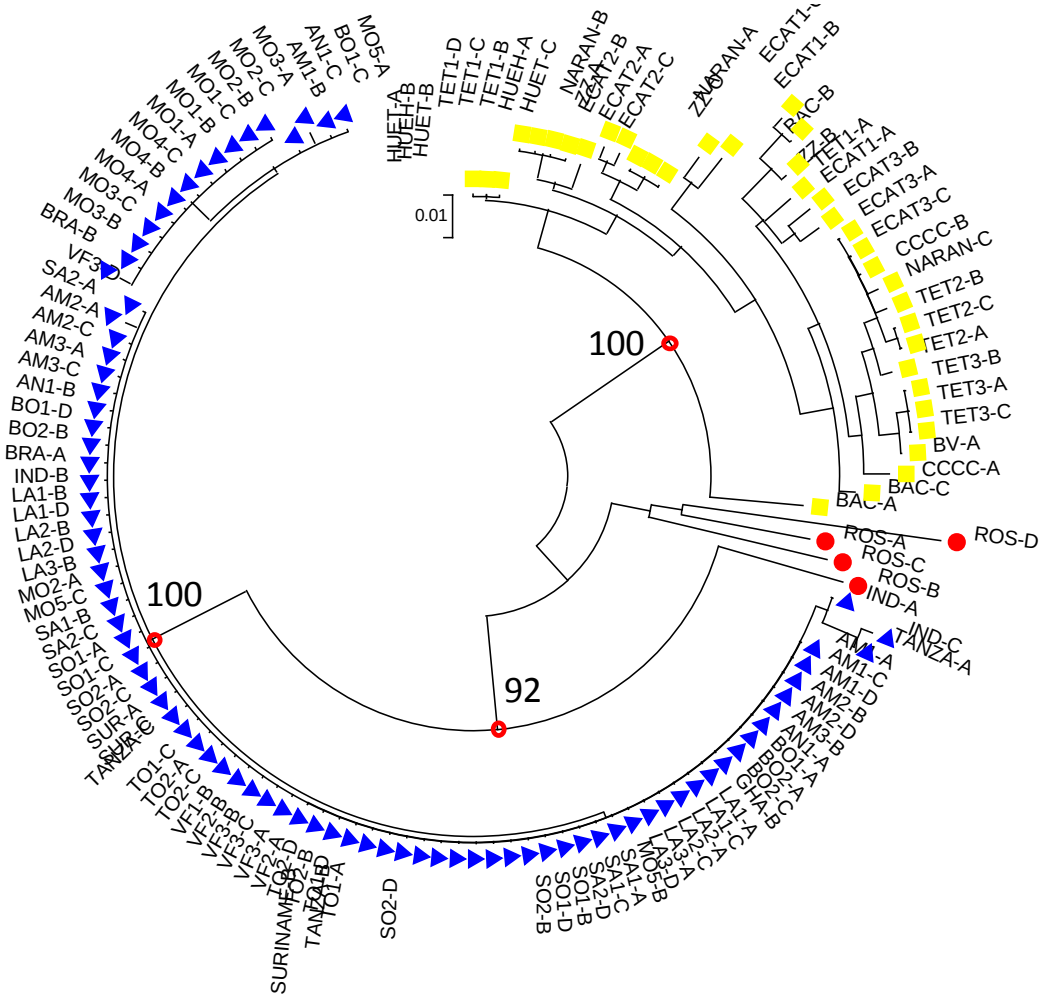


Figure 4.3 Midpoint rooted NJ tree of AFLP analysis in *J. curcas*. Bootstrap values (for nodes in open red circles) calculated based on 1000 replicates are indicated**.** Yellow colour represents the edible Mexican samples; red colour shows the samples from Mexico Rosario Izapa; blue colour stands for the non-edible samples outside Mexico. Scale bar represents Nei's genetic distance (Nei, 1973). Refer to Table 2.1 in Chapter2 for the samples' names.

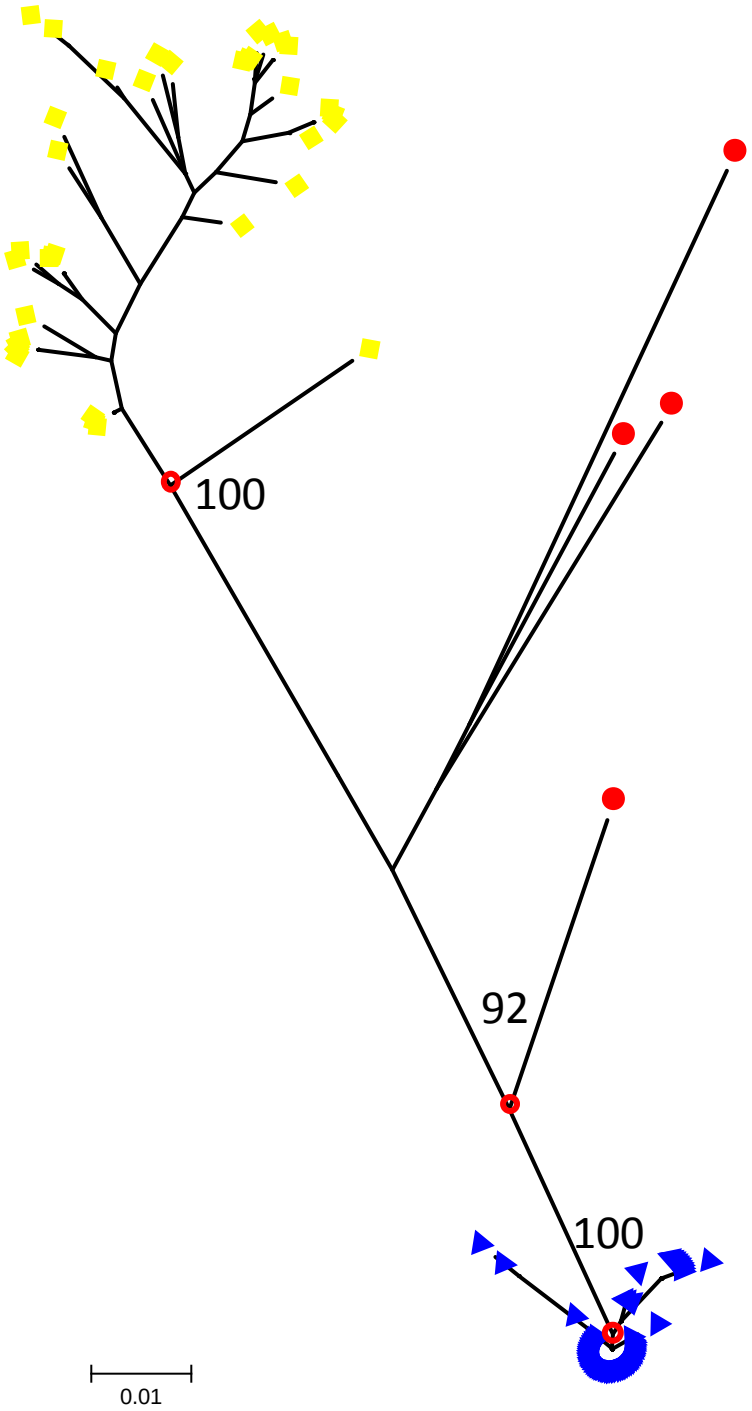


Figure 4.4 Unrooted radiation NJ tree of AFLP analysis in *J. curcas*. Sample names not shown. Bootstrap values (for nodes in open red circles) calculated based on 1000 replicates are indicated**.** Yellow colour represents the edible Mexican samples; red colour shows the samples from Mexico Rosario Izapa; blue colour stands for the non-edible samples outside Mexico. Scale bar represents Nei's genetic distance (Nei, 1973). Refer to Table 2.1 in Chapter2 for the samples' names.

4.4.2 Genetic structure analysis using Bayesian inference

Analysis using STRUCTURE software (Bayesian inference approach) was conducted to estimate the genetic structure in *J. curcas*. Based on the genetic variation observed in the phylogenetic trees, the entire collection of samples in this analysis were pre-defined as three "*subpopulations"* including edible, non-edible (excluding Rosario Izapa) and Rosario Izapa prior to the analysis. This "*subpopulation*" is a prior definition and therefore should not be confused with a real "population", which is conceptually equivalent to cluster (K value) revealed by the analysis. In Figure 4.5A, LnP (D) reaches a plateau when K=2 indicating that two clusters (populations) are sufficient to explain the data. Therefore there are two apparent populations in the collection of samples in this study.

In Figure 4.5B when K=2, the predefined edible and non-edible subpopulations are assigned into two clusters and exhibit substantial genetic homogeneity, apart from one sample from Bacalar, Quintana Roo in the edible subpopulation which shows ~10% probability towards the other cluster. Regarding the Rosario Izapa (R.I) subpopulation, they exhibit an admixture of the two clusters: R.I-B shows about 75% probability to the cluster which contains the non-edible subpopulation, whereas R.I-A, C and D show more than 50% probability to the cluster which contains the entire edible subpopulation. When K=2, most of the individuals (121/126) can be assigned with > 95% probability into one of the clusters.

An assignment probability with K=3 was also conducted (Figure 4.5C) and this shows a better resolution of the clustering. As with the scenario in K=2, each of the predefined edible and non-edible subpopulations group into one of the three clusters with remarkable genetic homogeneity. Moreover samples from Rosario Izapa are assigned into an individual cluster, with the exception that the R.I-B shows 50% probability to the cluster which contains the non-edible subpopulation. When K=3, 124 out of 126 individuals are assigned with > 95% probability into 3 clusters. It is noteworthy from the current work that samples from the R.I not only share markers from both edible and non-edible varieties, but also have some unique markers. Considering the results shown above, the Rosario Izapa samples appear to belong to

a distinct population, but analysis of only four individuals is not enough for a significant increase in the LnP (D) value between K=3 to K=2 to be detected.

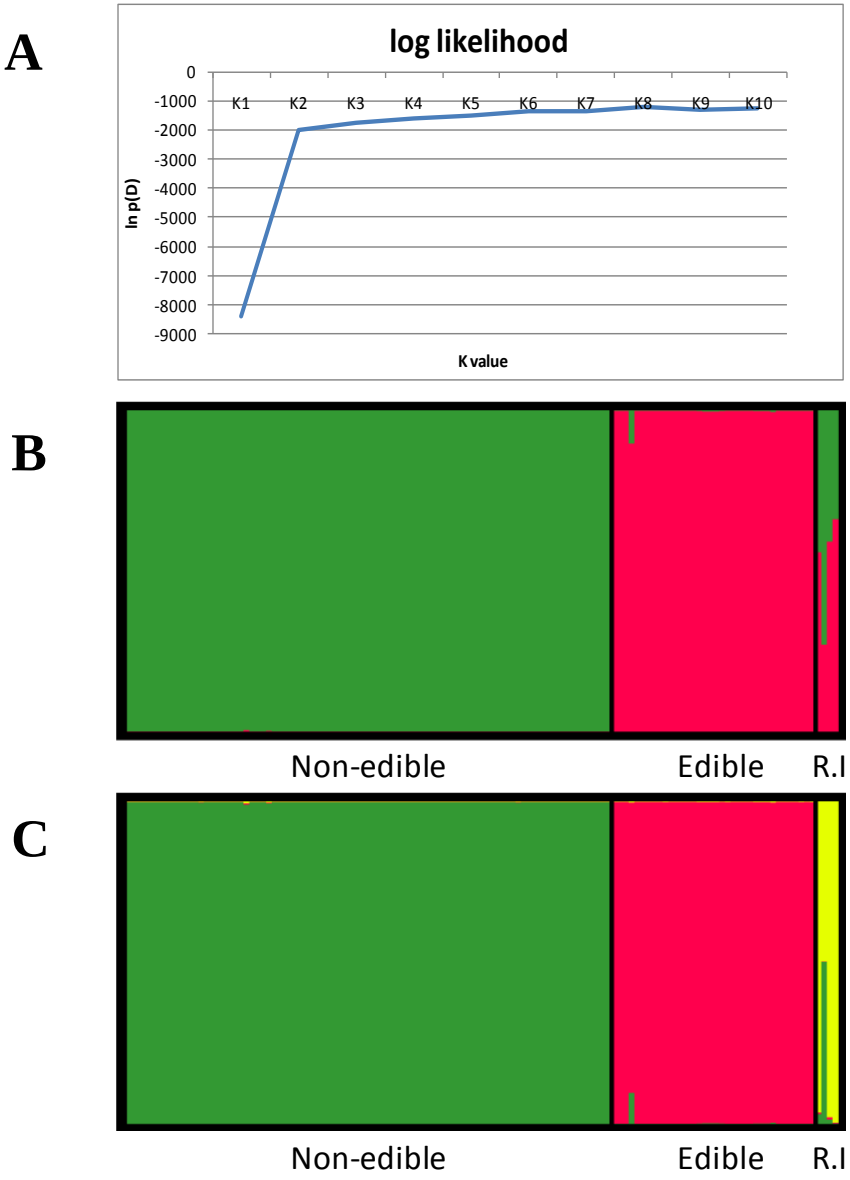


Figure 4.5 Analysis conducted using STRUCTURE (Bayesian analysis approach) to depict the genetic structure in *J. curcas*.

A: Log probability for cluster K= 1-10.
B and C: Individual assignment probabilities (model-based ancestry) of the collection of *J. curcas* samples under two and three clusters (K=2 in B and K=3 in C). Non-edible, edible and R.I (Rosario Izapa) are pre-defined subpopulations. Each colour represents one cluster. Each vertical line represents one individual and different colours represent a single cluster revealed by the analysis. The vertical height of each colour represents the probability of the individual within that cluster.

4.5 TaqI is a better frequent cutting restriction enzyme than MseI for *J. curcas* AFLP analysis

4.5.1 Distribution of amplicons

The distribution of multi-loci amplicons is important for gel scoring, especially when automated scoring is not available. Overlapping amplicons make gel scoring difficult and therefore legible amplicons with scattered distribution are always preferable. Figure 4.6 shows the size distributions of the fragments observed when AFLP analysis was performed using MseI and EcoRI. The LICOR system visualizes the fluorescently labelled amplicons ranging from 50 bp to 700 bp. With MseI primers, most of the amplicons detected were <300 bp and the fragment distribution was relatively narrow. The mean size of the fragments amplified is 209 bp, with a standard deviation of 124 bp. In contrast with MseI, a broader distribution of fragment sizes was observed using TaqI. The mean size of the fragments amplified is 285 bp, with a standard deviation of 163 bp.

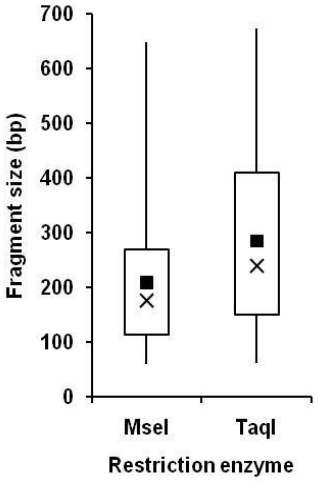


Figure 4.6 Distribution of fragment sizes observed when performing AFLP analysis with MseI and TaqI as frequent cutters. Vertical lines represent the size distribution of fragments observed between 50 and 700 bp. The boxed region represents fragment sizes between the upper and lower quartiles. The median is represented by an "x" and the mean by a closed square. For MseI, n = 146, and for Taq1, n = 191.

4.5.2 Discriminatory power

In any genetic marker studies, only the polymorphic markers are informative. PPB is the percentage of the polymorphic bands over the total observed bands and is a good indicator for utility of a particular primer pair. Due to the dominant nature of AFLP markers the PIC*i* value, which is the polymorphic information content of a certain marker *i*, peaks when the allele frequency (f*i*) is 50% [PIC*i* = 2f*i*(1-f*i*)]. A primer set with higher PPB does not guarantee a higher PIC, and *vice versa*. MI (marker index) considers both the number of polymorphic markers and PIC and thus is more comprehensive.

According to Table 4.3, on average, AFLP analysis using TaqI as the frequent cutter produced a higher number of polymorphic markers per primer pair than using MseI as the frequent cutter (on average, 10.2 in TaqI and 4.3 in MseI). The TaqI based AFLP analysis therefore resulted in a higher MI than MseI based AFLP analysis (on average, 2.47 in TaqI and 1.19 in MseI). The most effective primer combination is E-ACA/T-CG, which scores PPB of 35.9%, MI of 4.38.

Table 4.3 Polymorphism observed with different AFLP primer combinations.

Primer set	Polymorphic bands	Total bands	PPB	PIC Values		MI
				PIC_{av}	$PIC_{av(p)}$	
EcoRI/MseI primer combinations						
E-ACC/M-CAA	1	19	5.3%	0.022	0.427	0.43
E-AAC/M-CTT	5	30	16.7%	0.049	0.295	1.48
E-AAG/M-CTT	4	30	13.3%	0.041	0.307	1.23
E-AAG/M-CG	3	26	11.5%	0.045	0.392	1.18
E-AAC/M-CAG	4	21	19.0%	0.042	0.223	0.89
E-ACA/M-CG	9	20	45.0%	0.096	0.213	1.92
Minimum	1	19	5.3 %	0.022	0.213	0.43
Maximum	9	30	45.0 %	0.096	0.427	1.92
Average	4.3	24.3	18.5%	0.049	0.309	1.19
EcoRI/MseI total	26	146	17.8%			
*EcoRI/*Taq*I primer combinations*						
E-AAG/T-CC	6	34	17.6%	0.062	0.352	2.11
E-ACA/T-CAA	13	33	39.4%	0.093	0.236	3.07
E-ACA/T-CG	14	39	35.9%	0.112	0.313	4.38
E-AAC/T-CT	12	43	27.9%	0.064	0.230	2.76
E-ACC/T-CG	12	24	50.0%	0.071	0.142	1.70
E-ACC/T-CAC	4	18	22.2%	0.044	0.197	0.79
Minimum	4	18	17.6%	0.044	0.142	0.79
Maximum	14	43	50.0%	0.112	0.352	4.38
Average	10.2	31.8	32.2%	0.074	0.245	2.47
*EcoRI/*Taq*I totals*	61	191	31.9%			
Grand total	87	337	25.8%			

4.6 Discussion

4.6.1 Genetic diversity and genetic structure in J. curcas

The AFLP analysis shows that there is remarkably low genetic diversity in the samples from Madagascar, Brazil, India, Tanzania, Ghana and Suriname (PPB=3.3%), despite the obvious geographical isolation of these samples. The genetic diversity within the samples from Madagascar is even lower (PPB=1.5%). More genetic variation exists in the Mexican edible samples (PPB=8.9%), and the high genetic variation observed in the Mexican non-edible Rosario Izapa (R.I, PPB=17.2%) samples brings the PPB to 25.2% in the Mexican material. This high level of genetic diversity in Rosario Izapa also contributes significantly to the high genetic diversity level in the non-edible samples (PPB=20.5%), compared to the edible samples (PPB=8.9%).

According to both the genetic distance and Bayesian inference analyses, there is a remarkable distinction at the genetic level between the Mexican edible samples and the non-edible samples from outside Mexico. They separate into two clusters in the phylogenetic tree and are assigned into different populations according to the individual assignment probability. The Mexican Rosario Izapa samples are distinct from either of those two groups, but more analysis is needed to fully verify this observation.

These findings are consistent with the hypothesis that *J. curcas* is native to Meso-America: more natural variation is typically found in populations or accessions collected close to the geographic origin of a species (Heller, 1996). The AFLP analysis indicated that the edible samples from the states of Quintana Roo, Puebla and Vera Cruz are genetically similar. Schmook and Serralta-Peraza (1997) proposed that the edible varieties found in Quintana Roo may have been transported from other regions of Mexico, including Vera Cruz, Chiapas, and Yucatan by the Mayans and other settlers from 1970. Notably, the state of Puebla is next to the state of Vera Cruz, suggesting that the edible samples analysed in the current study may share a common ancestor.

The low genetic diversity in the samples from Madagascar, Brazil, India, Tanzania, Ghana, and Suriname indicates that all these non-edible samples might have the same origin. It has been suggested that *J. curcas* was transported from the Meso-America region via Cape Verde to Africa and Asia by the Portuguese sailors in the 16th Century (Heller, 1996). Similarly, due to the very low genetic diversity across Brazil *J. curcas* may have been introduced here also (Rosado et al., 2010). The similarity of the Surinamese sample to the other non-edible samples outside Mexico is likely to be a result of these seeds being sourced from outside Suriname, as no polymorphisms were observed within any of the three plants analysed from this batch of seeds. The re-introduction of non-native samples to the Americas is already documented; seeds from Cape Verde have been planted in Nicaragua (Foidl et al., 1996). In addition, the vegetative propagation *J. curcas* can also contribute to the limited genetic variation in certain regions/countries where it has been introduced.

The current AFLP result is in agreement with the low genetic variation which has been investigated in the germplasm from Asia (China, India, Sri Lanka, Jordan, Philippines, Indonesia and Thailand), Americas (Brazil) and Africa (Madagascar, Tanzania, Cape Verde, Uganda and Togo) (Sun et al., 2008, Ambrosi et al., 2010, Rosado et al., 2010, Sato et al., 2011, Shen et al., 2010). Though a few genetic marker techniques have been applied on a large collection of materials in India, it is not yet clear at which level the genetic variation exists, given the different levels of genetic variation reported in material from India (Basha and Sujatha, 2007, Sudheer et al., 2010a, Sudheer et al., 2010b, Yadav et al., 2010). Sudheer (2010b) suggested that different events of introduction of distant *J. curcas* germplasm to India may result in the consequent wide spread of different pedigrees. However, some of the results reported in the literature cited above are not convincing, because either RAPD markers were employed or data analysis lacked satisfactory statistical support, such as lack of the Bootstrap values in the phylogenetic trees. RAPD markers are susceptible to many interacting variables, such as DNA concentration, PCR conditions and even thermocycler equipment used (Weising et al., 2005). The irreproducibility of RAPD with an error rate of up to 25% has been reported by Jones et al (1997) and has been reviewed by Weising et al (2005).

In addition, a few studies have also compared the edible and non-edible varieties by various marker systems, and shown that these are distinct from each other (Sujatha et al., 2005, Basha et al., 2009, Sudheer et al., 2010a). Those studies have also shown that more genetic variation exists in the Mexican regions. Again, my results are in concordance with those observations.

In summary, results from the present AFLP study are consistent with the expectation. Given the low genetic variation outside Meso-America, caution should therefore be exerted before commencing any plant breeding programmes using materials sourced exclusively from those areas. There is more genetic variation in Mexico, especially in the state of Chiapas than in material collected from anywhere outside Meso-America. Therefore any further genetic diversity studies, or population selection effort should focus on materials from the Meso-America regions.

4.6.2 Higher performance of TaqI compared to MseI as frequent cutting restriction enzymes for J. curcas AFLP analysis

As a dominant marker system, AFLP has its competitive edge in terms of generating numerous, genome-wide di-allelic loci (Meudt and Clarke, 2007), and therefore more polymorphic bands would provide more informative capacity for any further analysis.

In this study, use of TaqI resulted in a significant improvement in the amount of polymorphisms detected per primer pair. It has been demonstrated that using TaqI as the frequent cutter in AFLP studies in animals leads to an increase of the portion of polymorphic bands (Bensch and Akesson, 2005). This can result from the underrepresentation of CpG dinucelotide in animal genomes (Beutler et al., 1989). The CpG dinucleotide is a site of methylation in animals (5-methylcytosine), and the majority of mutations that occur at this site are C → T and G → A transitions on the complementary DNA strand (Cooper and Youssoufian, 1988). Cooper and Youssoufian (1988) indicated those transitions are caused by methylation-induced deamination of 5-methyl cytosine. Similar to the animal genomes, in plants the CpG (and CpNpG) dinucletotide is also a methylation site (Finnegan et al., 1998) and could therefore also represent a mutational hotspot. A recent study of spontaneous mutation rates in *A. thaliana* has shown that G:C → A:T transitions are by far the

most frequent point mutation, as a result of two main processes: deamination of methylated cytosines and ultraviolet light–induced mutagenesis (Ossowski et al., 2010). Consequently it is supposed that in the AFLP studies of eukaryotes the use of TaqI (recognition site TCGA) as the frequent cutter would reduce the total number of amplicons and identify more mutations in the genome.

As well as capturing about 2-fold more polymorphisms, using TaqI as the frequent cutter the distribution of the amplicons on the gels is better than with MseI. MseI produces many bands clustered together at the bottom of the gels, thus making scoring relatively difficult. This can be attributed to the fact that TaqI might cut less frequently than MseI, due to a low GC content reported in the *J. curcas* genome [38.7% (Carvalho et al., 2008) and 34.3% (Sato et al., 2011)]. Containing an equal amount of GC/AT in its recognition site, the relative cutting frequency of TaqI is less sensitive to variations of genome GC content than MseI (Figure 4.7).

Based on the above observation, TaqI is recommended as a better frequent cutter in the AFLP analysis of *J. curcas*. Further the use of TaqI as a frequent cutter may be more widely applicable for AFLP analysis in other plant species.

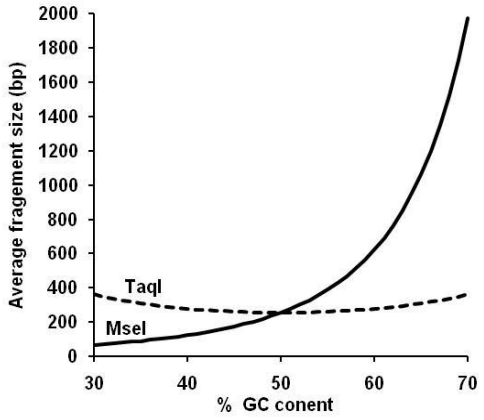


Figure 4.7 Effect of GC-content of genome on predicted mean restriction fragment size obtained with MseI (solid line) and TaqI (dashed line). The prediction is based on a random distribution of each base within the genome. For example, for the restriction enzyme TaqI, the fragment size for a 50% GC content is calculated as 1/ (0.25 *0.25*0.25*0.25).

Chapter 5: Characterization of curcin and induction of curcin by stress

5.1 Introduction

5.1.1 Ribosome Inactivating Proteins (RIPs) and curcin in J. curcas

Ribosome Inactivating Proteins (RIPs) are a heterogenous group of enzymes which possess N-glycosidase activity on ribosomal RNA and therefore are able to inhibit protein synthesis in eukaryotic cells. RIPs have been found in many plant species and also in a few bacteria and fungi (Barbieri et al., 1993, Stirpe and Battelli, 2006). The ribosome inactivating mechanism of RIPs have been widely studied, among which the observation that ricin (RIP from *Ricinus communis*) cleaves a single adenine base (A4324 in 28S rat rRNA), while α-sarcin (RIP from *Aspergillus giganteus*) cleaves the phosphodiester bond (between G4325 and A4326 in rat 28S rRNA), were extended to most of other RIPs (Stirpe et al., 1988). Apart from this classical enzymatic activity specifically toward the ribosome RNA, RIPs have proven to have polynucleotide: adenosine glycosidase activity (Barbieri et al., 1997). The substrates include DNA, poly (A) and also RNA. In addition, studies also indicate that individual RIPs may exhibit DNase (Li et al., 1991, Nicolas et al., 1997, Roncuzzi and Gasperi-Campani, 1996), RNase (Mock et al., 1996), phosphatase (Helmy et al., 1999), chitinase (Shih et al., 1997) and superoxide dismutase (SOD) (Sharma et al., 2004) activities. However some of these novel activities are still in dispute, challenged by problems of impurity in sample preparation (Peumans et al., 2001, Stirpe and Battelli, 2006).

To date three types of RIPs have been identified. Type I RIPs consist of a single enzymatic chain, which has an approximate molecular weight of 30 kD. Since RIPs are capable of inactivating protein synthesis, type I RIPs are highly toxic in cell free systems, but are much less toxic to the intact cell, as they can hardly penetrate the cell membrane. Type II RIPs are heterodimers which contain an enzymatic chain (A chain) and a carbohydrate binding lectin chain (B chain). The A chain is similar to type I RIPs, and the B chain helps the A chain enter the cell. As a result, type II RIPs exhibit ca. 1000-fold more cytotoxicity than type I RIPs (Barbieri et al., 1993). The two chains in type II RIPs are encoded by the same gene, and the approximate

molecular weight of type II RIPs is 60 kD. Type III RIPs are single chains comprising an enzymatic domain linked non-covalently to an extended polypeptide with unknown function. Type I and type II RIPs have been found in many plant species and also in a few bacteria and fungi (Stirpe, 2004); type III RIPs have only been found in maize (Walsh et al., 1991) and barley (Reinbothe et al., 1994b).

The role of RIPs in plants has been linked to defence against herbivores, insects and viruses as demonstrated *in vitro* and through the use of transgenic plants (Nielsen and Boston, 2001, Peumans et al., 2001). Due to their high toxicity, type II RIPs, such as ricin and abrin, are believed to protect the seed of castor bean (*Ricinus communis*) and jequirity peas (*Abrus precatorius*) against plant-eating organisms. Many type I RIPs have been found in various plants with proven antiviral activities toward plant viruses, such as pokeweed antiviral protein (PAP) from pokeweed (*Phytolacca acinosa*). It is still not clear how exactly RIPs render plants with resistance against viruses. Direct and indirect acting mechanisms have been proposed (Park et al., 2004). In the direct model, RIPs attack invasive pathogens, or trigger local cell suicide to avoid further infection. In the indirect model, they are assumed to mediate the activation of other defence systems in plants, possibly based on the fact that RIPs could disrupt protein synthesis (Park et al., 2004). However, no working mechanism or evidence has been shown to date.

In plants, gene families encoding several RIP isoforms exist (Hartley and Lord, 2004b). In *J. curcas*, Stirpe (1976) first isolated RIPs from the seed, and named it curcin. Subsequent studies identified and characterized two curcin genes, of which one is highly expressed in seeds (Lin et al., 2003a) and the other is expressed in leaves but only under stress conditions (Wei et al., 2005). These two genes are referred to as *CURCIN1* and *CURCIN2*. In our lab, an additional curcin gene sequence (denoted as *CURCIN3*) was identified by 454 sequencing of cDNA derived from developing seeds (Andrew King, unpublished result).

Although curcin has been suggested not to be as potent as phorbol esters as a toxic agent in *J. curcas* seeds, it is still considered the foremost toxic phytochemical in the hydrophilic extract from seeds (Devappa et al., 2010a). Lin et al (2003a) reported that purified curcin from *J. curcas* seeds strongly inhibits the cell-free translation in

the rabbit reticulocyte lysate with an IC_{50} (half maximal inhibitory concentration) of 0.19 nmol L^{-1}. The oral feeding and intraperitoneal injection of the purified curcin from *J. curcas* seeds into mice have shown an LD_{50} (median lethal dose) of 104.7 ± 29.4 mg kg^{-1} and 67.2 ±10.4 mg kg^{-1} respectively, suggesting that curcin is an acute toxin (Lin et al., 2010). Therefore, for any large scale production of feedstock using seed meal after oil extraction, eliminating curcin in the seeds is desirable. Since *J. curcas* seeds contain a number of other antinutrient proteins such as protease inhibitors, a heating process is necessary for the seed meal after oil extraction to be used as animal feed. However, as phorbol esters are non-detectable in the edible variety, it is interesting to determine if variation exists in curcin abundance in the edible variety as well as to compare the curcin level in the edible and non-edible varieties in *J. curcas*. Such studies have not been performed previously. If curcin is present in both edible and non-edible seeds, it would imply that heat treatment should be sufficient to inactivate this protein.

Like many other RIPs, curcin has antiviral and antifungal activities. Transgenic tobacco expressing *CURCIN2* has proven tolerant to tobacco mosaic virus (TMV) and a fungal pathogen *Rhizoctonia solani* (Huang et al., 2008), and Qin (2010) suggested CURCIN2 has a higher antifungal activity than CURCIN1 based on heterologous expression of both the curcins in *E.coli*. Therefore in spite of the interest to suppress curcin1 in seeds, at the same time it might be possible and favourable to retain curcin2 in the vegetative tissues of *J. curcas* to equip the plant with more tolerance to pathogens, herbivores and insects. I was therefore interested to determine the spatial and temporal expression pattern of the curcin gene family in *J. curcas*. This would lead to a better understanding of the tissue specificity of the different curcin genes.

5.1.2 The role of jasmonate in the regulation of curcin upon stress

In many plant species, RIPs can be induced under a variety of biotic and abiotic stress conditions, such as heat, osmotic stresses and senescence (Stirpe et al., 1996), salt-shock (Rippmann et al., 1997) and fungal infection (Xu et al., 2007). Wounding, jasmonic acid (JA) and abscisic acid (ABA) induced PIP2 in *Phytolacca insularis* (Song et al., 2000), and viral infection, H_2O_2 and salicylic acid (SA) induced two

RIPs from *Beta vulgaris* (Iglesias et al., 2005). The above elicitors such as JA, ABA and SA highlighted the possibility that the diverse expression patterns of RIPs under stress conditions are governed by phytohormones.

Phytohormones play important roles in regulating plant development and responses to biotic and abiotic stress (Bari and Jones, 2009). Among these phytohormones, jasmonate (JA), an oxylipin, together with its precursor 12-oxo-phytodienoic acid (OPDA), and its derivates methyl jasmonate (MeJA) and jasmonoyl-L-isoleucine (JA-Ile) (collectively referred to as jasmonates), are important plant defence gene regulators. In *Arabidopsis*, JA synthesis initiates in the plastids, where LOXs (LIPOXYGENASES) oxygenate tri-unsaturated fatty acids (18:3 and 16:3), to produce OPDA and dinor-OPDA, which are exported to the peroxisome. LOX2 is involved in the production of the majority of JA (Acosta and Farmer, 2010). In peroxisomes, OPDA and dinor-OPDA are reduced by OPR3 (OXOPHYTODIENOIC ACID REDUCTASE3), and generate JA by beta-oxidation. JA then serves as the substrate to form different derivatives such as JA-Ile which are biologically active (Acosta and Farmer, 2010). Jasmonates (JAs) mediate the stress response to insect-driven wounding, various pathogens, drought, low temperature and salinity (Wasternack and Parthier, 1997). The expression of an estimated 67-85% of wound- and insect-regulated genes are controlled by jasmonates in *Arabidopsis* leaves (Acosta and Farmer, 2010). In barley, both the endogenous methyl jasmonate (MeJA) level and JIP60 (a ribosome-inactivating protein) were induced upon wounding, and exogenously applied MeJA induced JIP60 (Reinbothe et al., 1994b). This suggested that jasmonates act as the intermediate to induce JIP60 (Reinbothe et al., 1994a).

In *J. curcas*, stress induction of curcin2 has previously been studied. Wei et al (2005) investigated how curcin2 levels change under a few biotic and abiotic stresses, including fungal infection, temperature (low and high) and PEG-6000 (Poly ethylene glycol, which mimic drought stress). The results showed that curcin2 is induced at both the transcript and protein levels under these stress conditions, and PEG treatment resulted in the most significant increase in curcin2 abundance. Qin et al (2009b) extended this study by treating the plants with more elicitors, such as UV light, ABA and SA, and showed that these stress conditions all induced curcin2 at

the transcript level to a certain extent. Taken together, it is interesting to study how curcin accumulates in response to other stress conditions, such as wounding, salinity and JA. It was expected that these treatments would induce curcin2, and therefore it was also interesting to establish if JA plays a role in the curcin regulation in *J. curcas* under stress conditions. Such work should contribute to understanding the regulation of RIPs in plants.

5.1.3 Aims

The aims of this study were to:

1. Compare the protein abundance of curcin in the dry seed of edible and non-edible *J. curcas* varieties.
2. Establish the spatial and temporal gene expression pattern and protein accumulation of the curcin gene family in *J. curcas*.
3. Investigate the regulation of curcin gene expression and protein accumulation under different environmental stresses.
4. Test the role of jasmonate in the regulation of curcin in response to various stresses.

5.2 Characterization of the curcin gene family

5.2.1 Identification of four curcin gene clones in J. curcas

In order to clone the curcin genes from *J. curcas*, both the edible variety from Mexico and the non-edible variety from Tanzania were used. PCR primers were designed based on the 454 database generated previously in our lab. As part of the present study an initial attempt to clone *CURCIN3* in the non-edible variety revealed another curcin gene (denoted as *CURCIN4*). All four curcin genes were cloned and sequenced from both cDNA and gDNA templates (Table 5.1). The primer pair designed for *CURCIN1* also amplified *CURCIN2* in the PCR reaction, because these primers anneal at sequences identical to *CURCIN1* and *CURCIN2*. However, using cDNA template, *CURCIN1* could only be amplified from the dry seed, whereas *CURCIN2* was cloned from the leaf. A single nucleotide polymorphism (SNP) was identified between the edible and non-edible *J. curcas* varieties in *CURCIN1*. At the

17 bp position of the open reading frame (ORF), a point mutation was observed (T in the non-edible Tanzania sample and C in the edible Mexican sample). This leads to a change of the amino acid from Met to Thr. *CURCIN3* was cloned from only the edible Mexican sample, and it was amplified in both leaves and dry seeds. Using the same primers, *CURCIN4* was identified in the non-edible Tanzania sample only and it was only amplified in the leaf. The genomic DNA sequences of the *J. curcas* curcin genes can be found in Appendix 4. It is worth mentioning that in my attempts to clone the above curcin genes, some sequences were obtained with high degree of similarity to the four curcin genes I cloned. However, when the experiments were repeated these sequences could not be verified. Thus they were excluded from any further investigation.

Table 5.1 Primers and DNA templates used for the successful cloning of the four curcin gene sequences. Samples from both edible (E) and non-edible (N) varieties were used as indicated in parentheses. cDNA was synthesized using RNA from dry seed and young leaf material. gDNA was extracted using young leaf material.

Genes	Primers	Template
CURCIN1	Curcin1 F1 / Curcin1 R2	seed cDNA (E/N) / gDNA (E/N)
CURCIN2	Curcin1 F1 / Curcin1 R2	leaf cDNA (E/N) / gDNA (E/N)
CURCIN3	Curcin3 F3 / Curcin3 R2	seed & leaf cDNA (E) / gDNA (E)
CURCIN4	Curcin3 F3 / Curcin3 R2	leaf cDNA (N) / gDNA (N)

Considering the length of the ORF region only, *CURCIN1* has a length of 879 bp, *CURCIN 3* has a length of 909 bp, and both *CURCIN2* and *CURCIN 4* are 927 bp long. An intron was found in the 5'-UTR region of each curcin gene by comparing the cDNA sequence with the gDNA sequence, and also by *in silico* intron prediction using NetGene2 (Brunak et al., 1991, Hebsgaard et al., 1996). All introns are located 12 bp upstream of the ATG initiation codon. The intron in *CURCIN1* and *CURCIN2* is 228 bp long, whereas *CURCIN3* and *CURCIN4* have introns of 223 bp. There are two polyadenylation signals (AATAAA) in the 3'-UTR of each curcin gene, which is typical for RIPs (Hartley and Lord, 2004b). A schematic representation of the structure of the curcin genes 1-4 is shown in Figure 5.1.

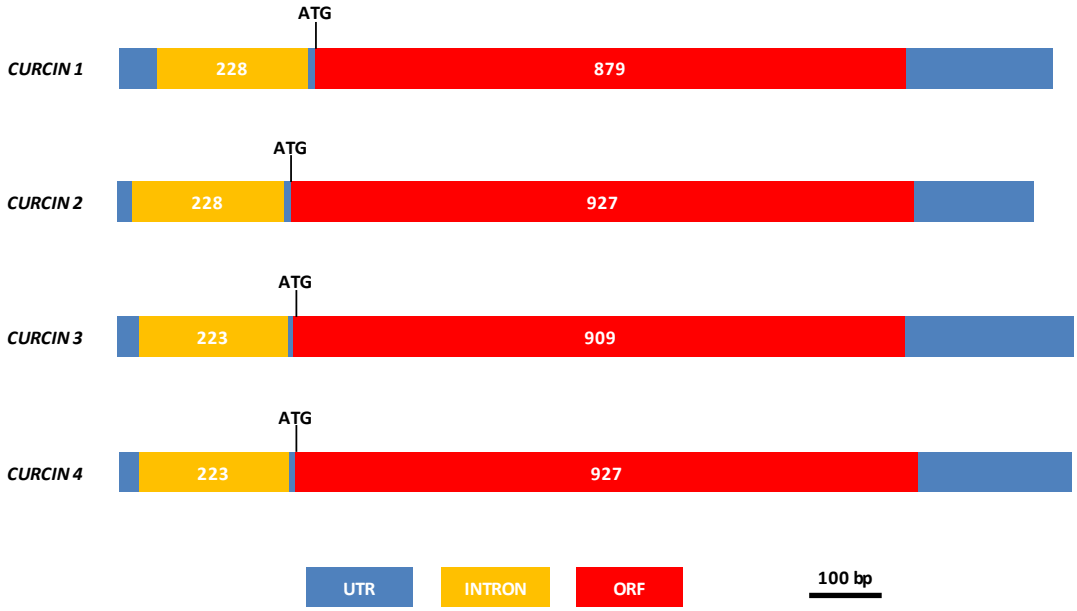


Figure 5.1 Gene structure of the four curcin genes in *J. curcas*. UTR, intron, and the ORF were shown in blue, yellow and red respectively. Numbers represent the length (bp) of the intron and ORF. Scale bar represents 100 bp.

The four curcin genes are markedly similar to each other (Figure 5.2). *CURCIN1* and *CURCIN2* share 94% identity, while *CURCIN3* and *CURCIN4* are 93% identical to each other (Table 5.2).

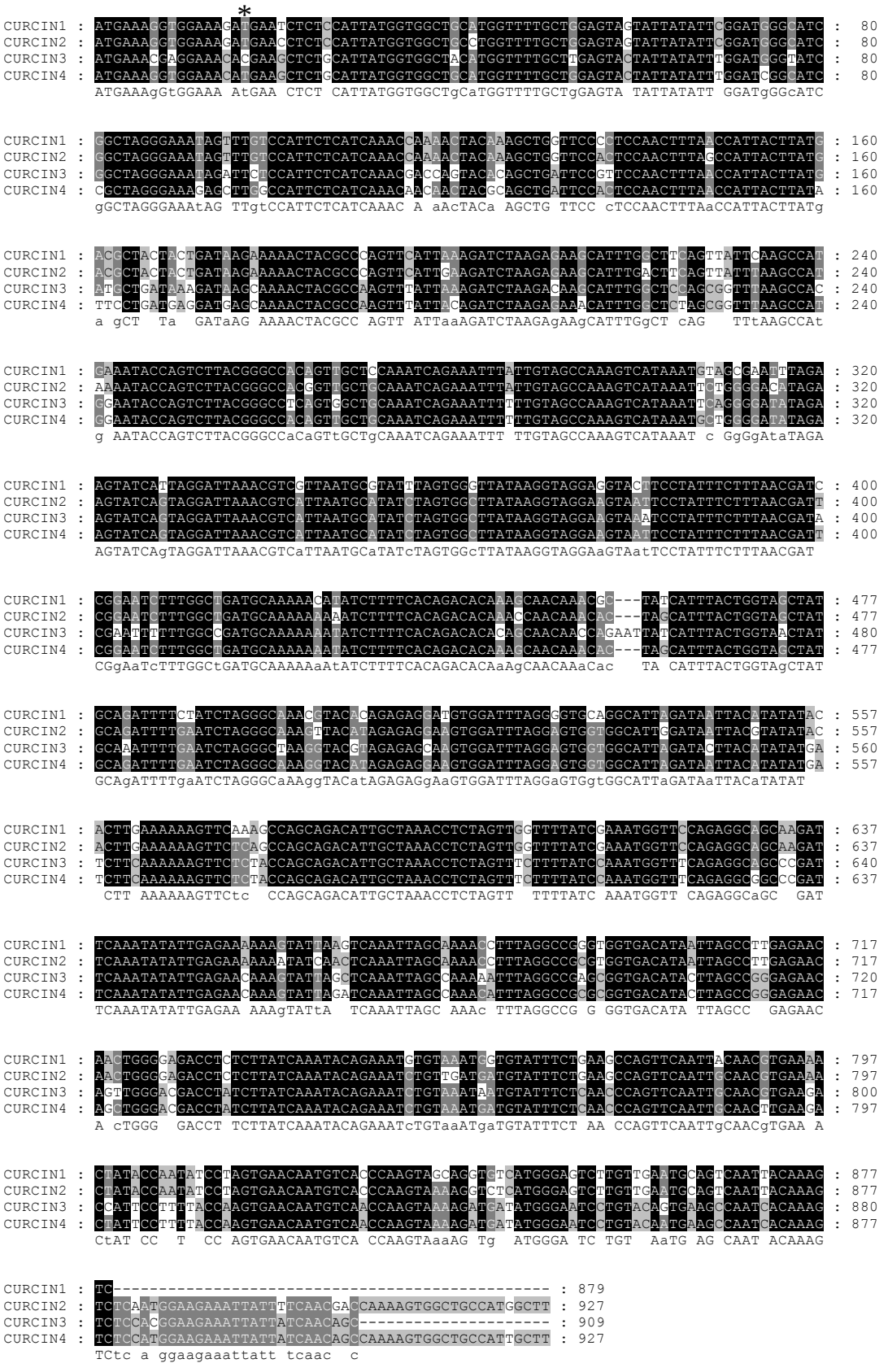


Figure 5.2 DNA sequence alignment of the ORF region of the four curcin genes. Alignment was performed using GeneDoc version 2.7 with "Identity" shading mode; the consensus sequence is indicated below the alignment. The asterisk shows the position of the SNP in *CURCIN1* between the non-edible and edible varieties.

Table 5.2 Identity between the four *J. curcas* curcin genes. The analysis compared the ORF for each individual gene. The % identity was calculated with AlignX integrated in the Vector NTI 10 program. Identity of the protein sequence is indicated in parentheses.

	CURCIN1	*CURCIN2*	*CURCIN3*	*CURCIN4*
CURCIN1		94 (87)	84 (71)	86 (74)
CURCIN2			86 (73)	88 (79)
CURCIN3				93 (86)
CURCIN4				

5.2.2 In silico analysis of curcin reveals that they are a highly conserved gene family

To better understand the properties of the four curcin proteins, their amino acid sequences were deduced from their DNA sequences. Each of the four curcin proteins has about 300 amino acids, and the theoretical molecular weight (Mw) is about 33 kD (Table 5.3). A predicted signal peptide with 28 residues was identified in each of the four curcin proteins by SignalP (http://www.cbs.dtu.dk/services/SignalP/), reducing their Mw by about 3 kD. Prediction of the pI values shows that curcin1 is a basic protein, all the other three curcins are acidic proteins. An alignment of the four curcin proteins is shown in Figure 5.3.

Table 5.3 Summary of the curcin proteins. The amino-acid sequences were deduced from the *CURCIN* genes. Signal peptide prediction was performed using SignalP (http://www.cbs.dtu.dk/services/SignalP/). Mw and pI were computed using Compute Mw/pI tool (http://expasy.org/tools/pi_tool.html). Properties of the mature protein following removal of the predicted signal peptide are given in parentheses**.**

Proteins	Deduced length	Mw	pI
	amino acid residues	kD	pH
CURCIN1	293 (265)	32.6 (29.5)	8.75 (8.36)
CURCIN2	309 (281)	34.9 (31.8)	6.20 (5.58)
CURCIN3	303 (275)	34.2 (31.0)	6.33 (5.49)
CURCIN4	309 (281)	34.9 (31.9)	4.95 (4.80)

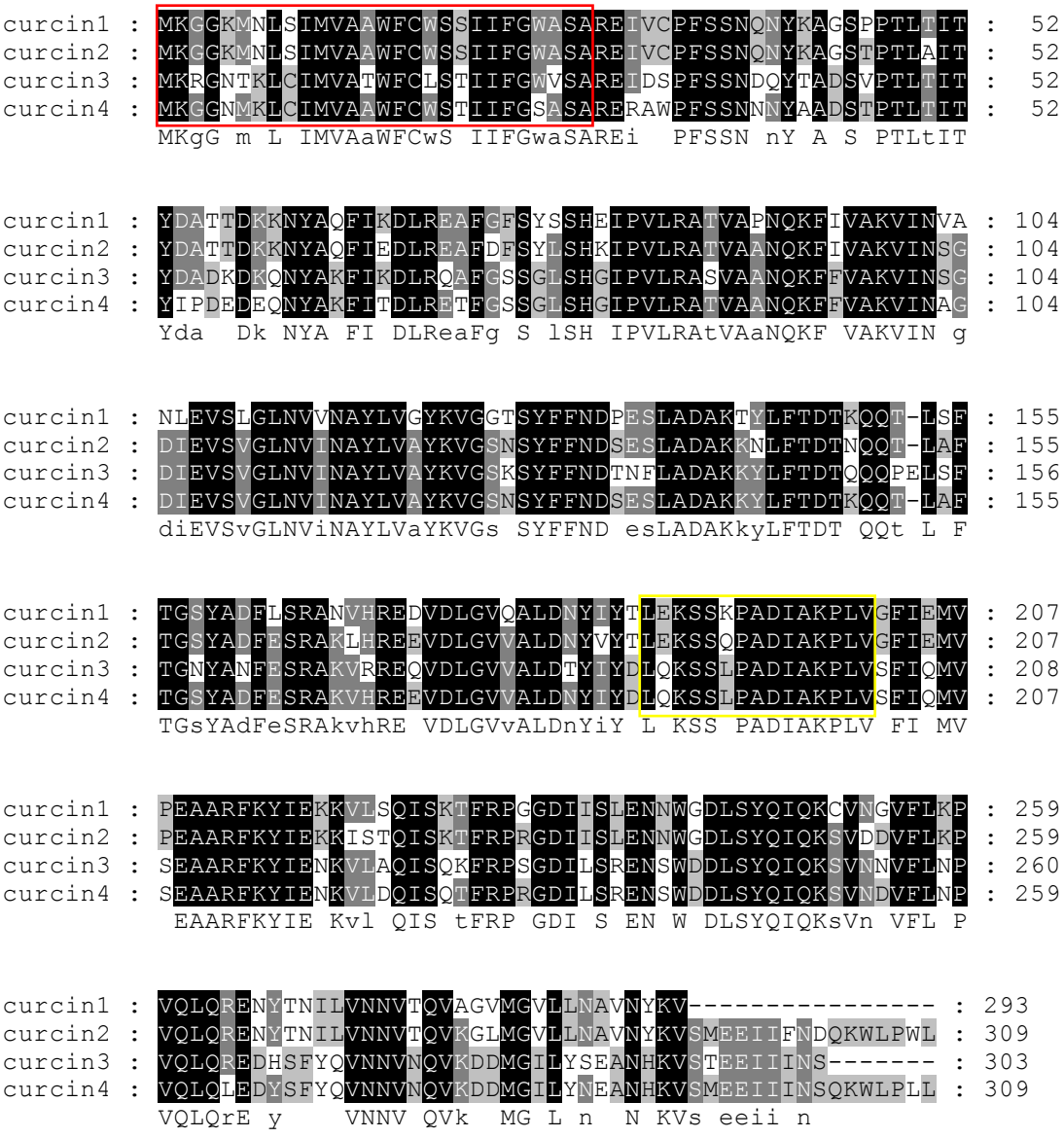


Figure 5.3 Alignment of the deduced amino acids sequences of the four curcin proteins. Alignment was performed using GeneDoc version 2.7 with "Tertiary" shading mode; the consensus sequence is indicated below the alignment. The signal peptide sequences were predicted using SignalP (http://www.cbs.dtu.dk/services/SignalP/), and are shown in the red rectangular box (28 residues). Yellow rectangular box shows the amino acids selected as the epitope for designing the anti-curcin peptide.

The protein sequences of the four curcins were compared with 6 other RIPs sequences previously deposited in GenBank and Swiss-Prot. Sequence alignment shows that twenty-two amino acid residues are highly conserved in all proteins (Figure 5.4). It has been previously suggested that five amino acid residues play an important role in the enzymatic activity of RIPs (Hartley and Lord, 2004a). These five residues are Tyr80, Tyr123, Glu177, Arg180 and Trp211 [RTA (ricin A chain) numbering], which are maintained in all four *J. curcas* curcin proteins.

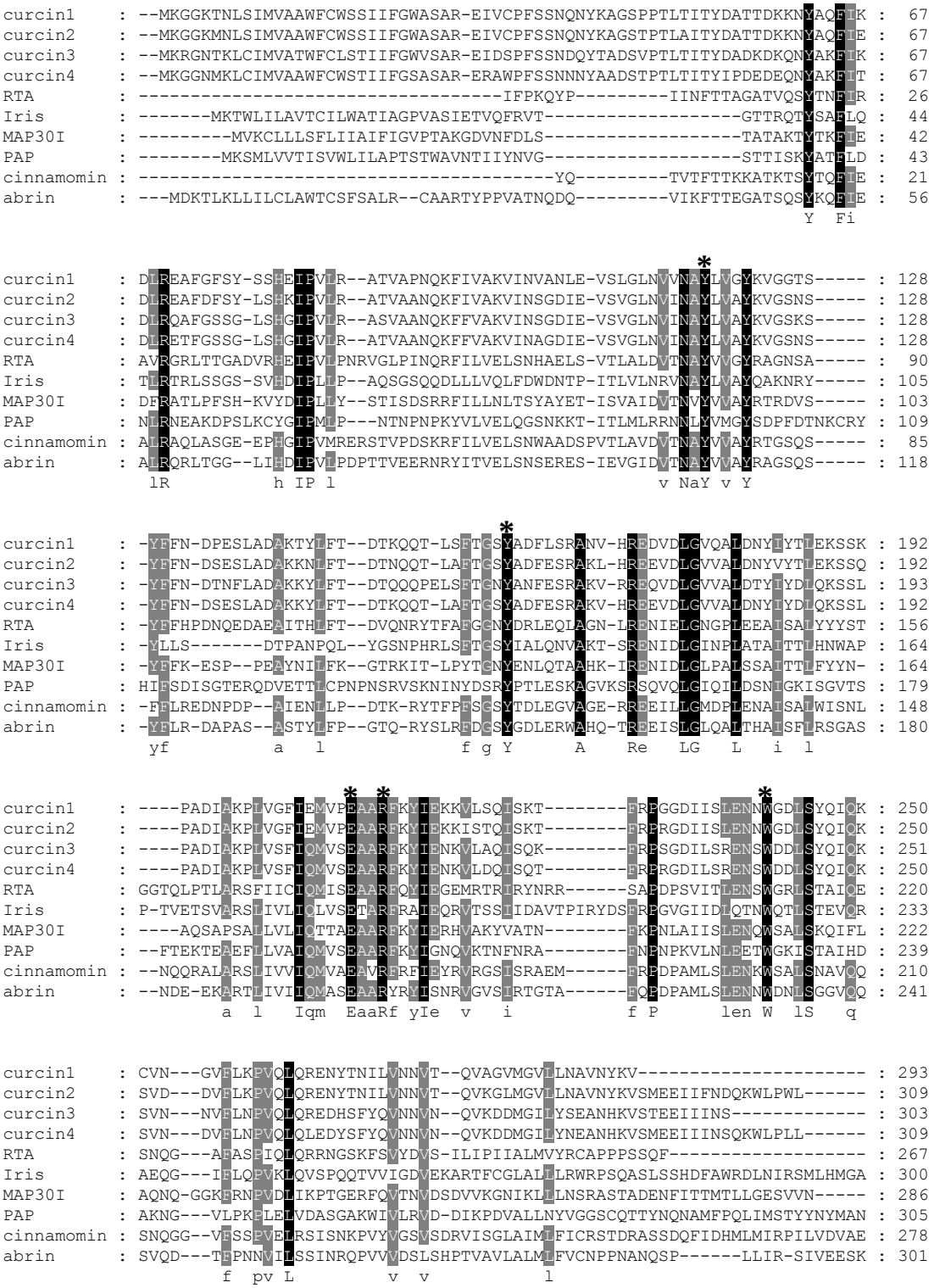


Figure 5.4 Comparison of the amino acids sequences of the four curcin with six other RIPs. [Protein sequences from GenBank and Swiss-Prot. Species names, numbers of amino acid residues used for the alignment, and accession numbers are indicated. RTA (ricin A chain, RIP from *Ricinus communis*, 1-267): P02879.1, Iris (RIP from *Iris* x *hollandica*, 1-300): AAC49780.1, MAP30I (RIP from *Momordica charantia*, 1-286): AAG33028.1, PAP (RIP from *Phytolacca acinosa*, 1-305): AAL15442.1, cinnamomin (RIP from *Cinnamomum camphora*, 1-278): AAF68978.2, abrin (RIP from *Abrus precatorius*, 1-301): CAA39202.1.]. Alignment was performed using GeneDoc version 2.7 with "Tertiary" shading mode. Conserved sites relevant to the *N*-glycosidase enzymatic activity are indicated by asterisks, including Tyr80, Tyr123, Glu177, Arg180 and Trp211 (RTA numbering).

The protein sequences of the four curcins were compared with those deposited in GenBank, and in recently published work (Sato et al., 2011, Yin et al., 2010) (Table 5.4). CURCIN1 differs from seed-specific curcin AY069946.1 by Lin et al (2003a), but is identical to the predicted protein of Sequence 1 and JcCA0072631.10. This indicates that PCR or sequencing errors might have been introduced in the sequence of AY069946.1, or it is possible that different alleles might exist in this CURCIN1 isoform. CURCIN2 is more consistent, including the predicted protein sequences from EU395775.1, EU195892.1, AY435214.1, Sequence 2 and JcCB0153511.10. The full length nucleotide sequences encoding CURCIN3 and CURCIN4 have been cloned only in this study, although were partially recovered by genome sequencing in Sato et al (2011). The low identity in their amino acid sequences might result from a few contig assembling errors. FJ357424.1 encodes a protein most similar to CURCIN1, and JF357910.1 and Sequence 3 encodes a protein which mostly resembles CURCIN2. Their relatively low identity to the curcin proteins in this study suggests that they might encode different curcin isoforms.

Table 5.4 Comparison of the four curcin in this study with others in GenBank and other studies. AA: amino acid. Yin et al see (Yin et al., 2010); Sato et al see (Sato et al., 2011).

Curcin identified in the current study	Nucleotide (Protein)	Identity of AA sequence	Length	Source
CURCIN1 (293 AA)	AY069946.1 (AAL58089.1)	98%	Full-length	GenBank
	Sequence 1	100%	Full-length	Yin et al
	JcCA0072631.10	100%	Full-length	Sato et al
CURCIN2 (309 AA)	EU395775.1 (ABZ04128.1)	100%	Full-length	GenBank
	EU195892.1 (ABW17545.1)	99%	Full-length	GenBank
	AY435214.1 (AAR08395.1)	99%	Full-length	GenBank
	Sequence 2	100%	Full-length	Yin et al
	JcCB0153511.10	100%	Full-length	Sato et al
CURCIN3 (303 AA)	JcCB0167631	93%	Partial	Sato et al
	JcCA0243151	93%	Partial	Sato et al
	JcCA0001311	95%	Partial	Sato et al
CURCIN4 (309 AA)	JcCB0761441.10	98%	Partial	Sato et al
	JcCD0114051.10	92%	Partial	Sato et al
	JcCA0001311	93%	Partial	Sato et al
Unknown	FJ357424.1 (ACO53803.1)	85% to CURCIN1	Full-length	GenBank
	JF357910.1 (AEA72440.1)	90% to CURCIN2	Full-length	GenBank
	Sequence 3	88% to CURCIN2	Full-length	Yin et al

5.2.3 Polyclonal antibody raised against a CURCIN1 peptide also detects CURCIN2

Previously reports have shown that CURCIN1 is abundant in the dry seed of *J. curcas* in the non-edible variety (Lin et al., 2003b, Qin et al., 2010). In addition another seed curcin gene was identified (*CURCIN3*). In order to compare the curcin level in the dry seed in the edible and non-edible *J. curcas* varieties, an antiserum against the peptide of a conserved region in the four *J. curcas* curcin proteins was produced (denoted as anti-curcin). The epitope chosen for the anti-peptide production is based on the amino acid residues from CURCIN1 (H_2N-LEKSSKPADIAKPLVC-COOH). In this epitope region, CURCIN2 has 1 residue different from the epitope, and both CURCIN3 and CURCIN4 have the same residues, which differ from the epitope by 2 residues (Figure 5.3, shown in the yellow rectangular box). Due to the high similarity of the four curcin proteins in the epitope region, this antibody might also cross-react with the other curcin proteins apart from CURCIN1. This would serve for the determination of the protein accumulation pattern of curcin, as described in the aims of this chapter.

The anti-curcin antibody might have different affinity with the four curcin proteins, because of the variation in the epitope region of the four curcin proteins (Figure 5.3, shown in the yellow rectangular box). In order to determine how well the anti-curcin antibody will react with the four curcin proteins, these were first expressed in *E. coli*, and the affinity of the anti-curcin antibody was tested. The ORFs of the four curcin genes were cloned into the pET28a vector containing an N-terminal hexahistidine epitope tag and expressed in *E. coli*. SDS-PAGE (Figure 5.5A) and immunoblot (Figure 5.5B) analysis using an antiserum specific for the hexahistidine epitope tag show that all four recombinant curcin proteins are in the insoluble fraction, possibly in inclusion bodies.

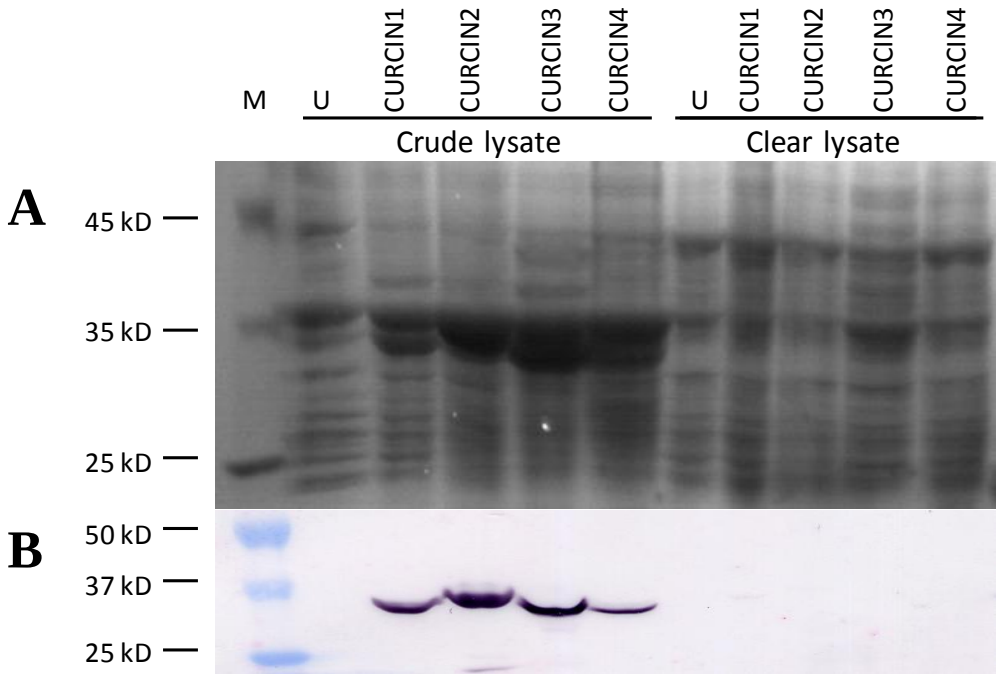


Figure 5.5 Heterologous expression of the four curcin proteins in *E. coli*. M: marker; Samples U and curcin 1-4: protein extract from *E. coli* cells containing either an empty pET28a vector (U), or a pET28a vector containing the construct of the ORF of a curcin gene with N-terminal hexahistidine tag. The crude lysate and clear lysate show the total proteins and the soluble proteins, respectively.
A: SDS-PAGE analysis of the heterologous expression of the four curcin proteins in *E. coli*.
B: Immunoblot analysis using anti-histidine of the heterologous expression of the four curcin proteins in *E. coli.*

To determine the specificity of the anti-curcin antibody against the four curcin proteins, the same amount of the crude cell extracts were denatured and loaded onto two gels (Figure 5.6A and B). Immunoblotting was performed using anti-histidine (Figure 5.6A) and anti-curcin antibodies (Figure 5.6B), respectively. The band density of that the anti-histidine antibody recognizes the heterologously expressed curcin proteins should reflect the actual amount of the four curcin proteins present in the samples loaded onto the gel (Figure 5.6A, CURCIN3> CURCIN2> CURCIN1> CURCIN4). In Figure 5.6B, the anti-curcin antibody recognized both CURCIN1 and CURCIN2 very well. In CURCIN2 an unspecific band was observed (the lower band). Because there was much less CURCIN1 than CURCIN2 in the sample, the anti-curcin antibody should have a better recognition of CURCIN1 than CURCIN2. On the other hand, CURCIN3 could barely be detected by the anti-curcin antibody in spite of an adequate loading; and CURCIN4 could not be detected at all. Since CURCIN3 and CURCIN4 share a common peptide sequence in the epitope region, the result suggests that CURCIN3 and CURCIN4 cannot be recognized by the anti-curcin antibody. The discriminatory affinity of the anti-curcin against the four

curcin proteins can be attributed to the variation of AA sequences in the epitope regions.

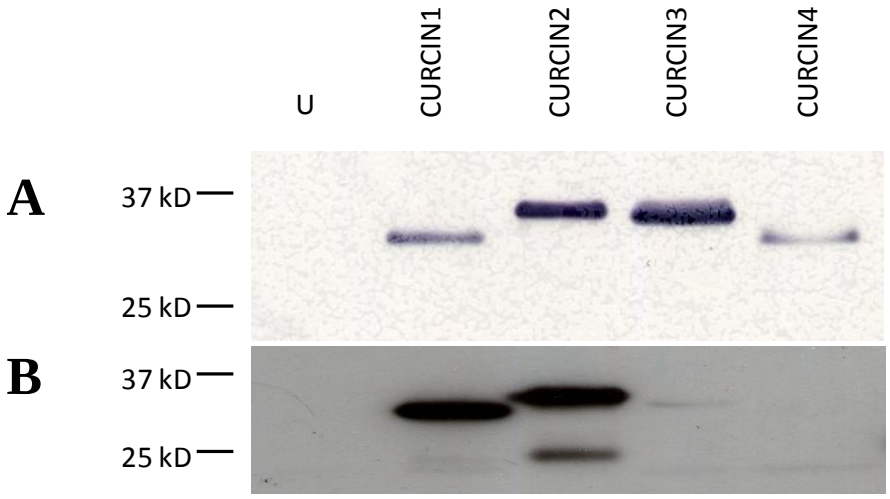


Figure 5.6 Determination of the anti-curcin antibody specificity with the heterologously expressed curcin proteins in *E. coli*. Proteins were extracted from the crude *E. coli* culture. The same amount of protein samples was loaded in gels A and B. Refer to Figure 5.5 for the details of the heterologous expression of the four curcin proteins in *E. coli*. U: expression of the empty vector; curcin 1-4: *E. coli* expressed curcin proteins. A: Immunoblot analysis using anti-histidine antibody with the heterologously expressed curcin proteins in *E. coli.* B: Immunoblot analysis using anti-curcin antibody with the heterologously expressed curcin proteins in *E. coli.*

.

5.3 CURCIN1 is predominantly endospermic, and accumulates at equal concentration in edible and non-edible seeds

Since no detectable phorbol esters exist in the edible seed variety (Chapter 3, Section 3.2.3), it is interesting to investigate whether CURCIN1, another toxic component of seeds, varies in its abundance in the edible and non-edible varieties (CURCIN3 might be present in the dry seed, but it cannot be recognized by the anti-curcin antibody). Thus this analysis was performed and the results are shown in Figure 5.7. In the SDS-PAGE, a band above 25 kD (Figure 5.7A, indicated in the red rectangular box) was recognized by the anti-curcin antibody in the immunoblot analysis (Figure 5.7B). This agrees with the expected size of curcin protein (Table 5.3). In the previous section we have demonstrated that the anti-curcin antibody strongly recognizes CURCIN1 (Figure 5.6B). In order to confirm that this band is indeed CURCIN1, it was subjected to protein identification by MALDI-TOF analysis. The MASCOT programme matched three tryptic peptides to the deduced peptide sequence of the CURCIN1 (Figure 5.7C). Hence, CURCIN1 levels in the edible and non-edible varieties are relatively equal (Figure 5.7A and B). I further examined the localization of CURCIN1 in the seed in *J. curcas*. Figure 5.8 shows that CURCIN1 is mostly present in the tegmen and endosperm, but its expression is very low or non-detectable in the embryo.

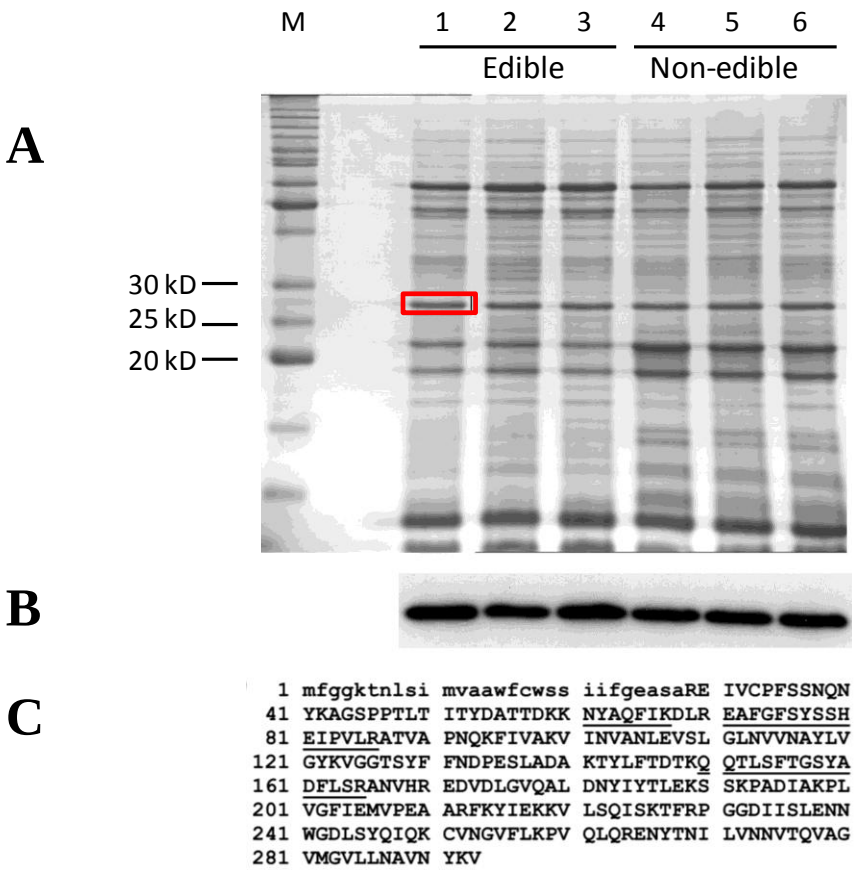


Figure 5.7 Analysis of relative CURCIN1 concentrations in seeds from the edible and non-edible varieties of *J. curcas*.
A and B: SDS-PAGE and immunoblot analysis of curcin1 containing extracts obtained from three Mexican edible *J. curcas* samples (1 - Bacalar, Quintana Roo; 2- Ecatlán, Puebla; and 3- Tetelilla-Puebla) and non-edible samples (4 - of unknown provenance, 5- Amparaky, Moyen Ouest, Madagascar; and 6 - Diligent of Tanzania). The band in the red rectangular box shows CURCIN1.
C: Deduced peptide sequence of CURCIN1 with the three tryptic peptides (underlined) identified by MALDI-TOF analysis. The prediction signal peptide identified by SignalP (http://www.cbs.dtu.dk/services/SignalP/) is denoted in lower case letters.

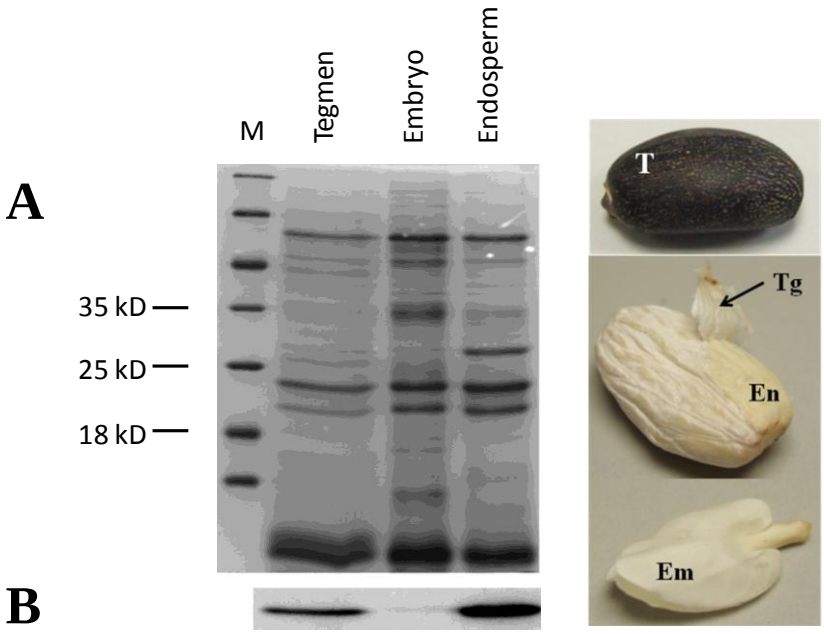


Figure 5.8 Analysis of CURCIN1 distribution within the seeds of *J. curcas* by (A) SDS-PAGE and (B) immunoblot analysis. T represents the testa. Protein extracts were obtained from the tegmen (Te), embryo (Em) or endosperm (En).

5.4 Differential temporal and spatial gene expression and protein accumulation patterns of curcins during plant development

5.4.1 Transcript levels of curcin in J. curcas by qPCR

To determine the spatial and temporal expression pattern of the four curcin genes in *J. curcas*, a time-course was established of development during seed germination and seedling establishment, and in the vegetative tissues (see Chapter 2, Figure 2.1). Samples were collected, and a quantitative, real-time (qRT) PCR analysis was performed as described in Chapter 2.

The expression pattern of the *J. curcas* curcin genes is shown in Figure 5.9 (*CURCIN3* expression could not be detected in the qPCR, thus it has been excluded from any further analyses). *CURCIN1* transcript is mostly endospermic; very low or no detectable *CURCIN1* expression was found in the other tissues (Figure 5.9A). In the endosperm, the transcript of *CURCIN1* is mainly present in dry seed, 2-3 DAI (days after imbibition) and WE (wizened endosperm) stages, peaking at 3 DAI.

CURCIN2 is expressed in the seedling, cotyledon, endosperm, and in all vegetative tissues (Figure 5.9B). During germination and seedling establishment, *CURCIN2* expression gradually increases from 2 DAI in the seedling, peaks at 6 DAI and decreases in the cotyledon in the WE stage; in the endosperm it is expressed from 2 DAI to the WE stage, with the highest abundance appearing at 3 DAI. In the vegetative tissues, its expression level in CS-cotyledon and TL-true leaf is higher than in TL-cotyledon, stem and root (10-50 fold more).

CURCIN4 is expressed in the seedling, cotyledon and all vegetative tissues, but not in the endosperm (Figure 5.9C). Its expression starts and peaks at 2 DAI in the seedling immediately after germination, decreases and maintains till 6 DAI. It should be noted that in Figure 5.9A, B and C, the magnitude of the scales of the y-axis differs. This is due to differences in the expression levels of these genes. For example, *CURCIN2* is much more abundant than *CURCIN1* and *CURCIN4* (100-500 fold more). The expression levels of *CURCIN2* and *CURCIN4* in vegetative tissues were compared. The expression level of *CURCIN4* in the cotyledon, true leaf, stem and root, though apparent in Figure 5.9C, is actually negligible compared to *CURCIN2* (Figure 5.9D) expression.

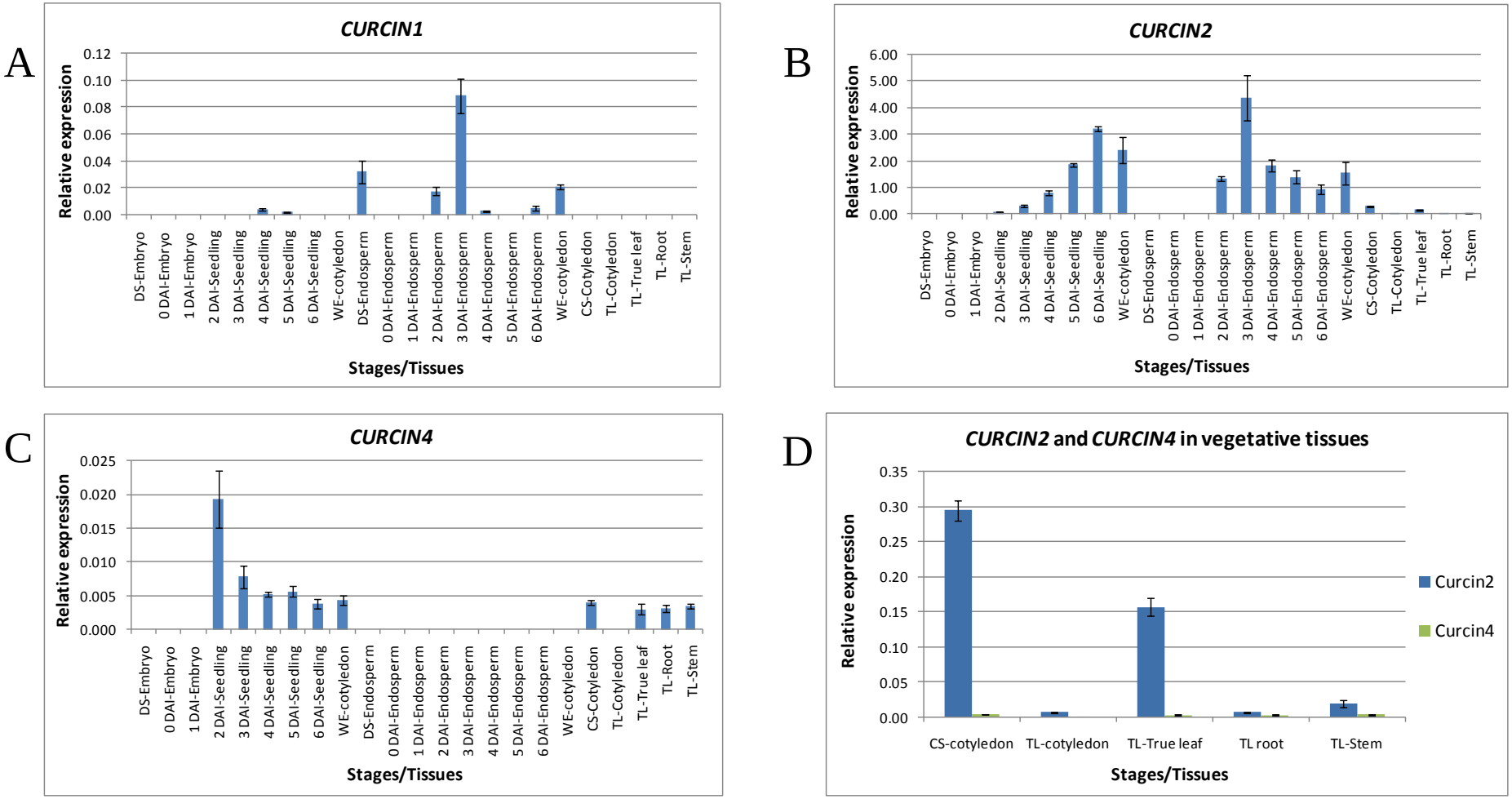


Figure 5.9 Expression of the curcin genes in *J. curcas* during seed germination and seedling establishment, and in the vegetative tissues by qRT-PCR. All the expression levels were normalized against *ACTIN*. The error bars represent standard error mean (SEM) of 3 technical replicates. DAI: days after imbibition; DS: dry seed; WE: wizened endosperm; CS: cotyledon spread; TL: true leaf. Refer to Chapter 2 for the details of the stages.

5.4.2 Time-course analysis of curcin protein abundance

To determine the pattern of curcin protein accumulation and compare it to the gene expression data, the following analysis was performed using the same material from the qRT-PCR analysis. Three bands of different molecular weights (band A, B and C) were detected by the anti-curcin antibody in the 23 samples analyzed during the time-course (Figure 5.10). It is worthwhile stating first that the accumulation of CURCIN1 protein occurs during seed development (Qin et al., 2009a). In Section 5.3 it has been demonstrated that CURCIN1 protein is abundant in the dry seed, in particular in the endosperm (Figure 5.7 and Figure 5.8).

Band C was identified as CURCIN1 based on its molecular weight being the same as the control sample (seed protein extract containing CURCIN1). CURCIN1 is present in the endosperm at a high level from the dry seed to the 6 DAI (Figure 5.10B), and even in the wizened endosperm (Figure 5.10C). It is most likely to be CURCIN1 protein already accumulated during seed development. CURCIN1 was also found in the embryo of the dry seed, but this might be a contamination from the endosperm material in the embryo sample due to the difficulty of adequately separating these two tissues. It is possible, but unlikely, that CURCIN1 has accumulated in the embryo at a very low level during the seed development.

Band A of a protein with a molecular weight of about 2-3 kD higher than CURCIN1 was found to accumulate from 3 DAI in the endosperm and 5 DAI in the seedling (Figure 5.10A and B). It is also present in the cotyledon in the WE, CS and TL stages but in a very low amount or undetected in the true leaf, stem and root. The molecular weight prediction (Table 5.3) has suggested that the mature protein of native CURCIN1 has a molecular weight of 29.5 kD and CURCIN2 is about 31.8 kD. In addition, according to Figure 5.6, heterologous CURCIN2 has a higher molecular weight than heterologous CURCIN1, whereas CURCIN3 and CURCIN4 could not be detected. Furthermore, the gene expression data revealed that *CURCIN2* expression strongly correlates with the accumulation of band A in the western blot analysis (compare Figure 5.9B with Figure 5.10).

Thus it is suggested that band A is CURCIN2 [this is also supported by the stress induction study (see Section 5.5.2)]. In the seedling, the expression of *CURCIN2* starts increasing from 3 DAI and it is in 5 DAI that CURCIN2 can be detected at the protein level. The accumulation of the CURCIN2 protein in the seedlings results in CURCIN2 being detected in the cotyledon samples in WE, CS and TL stages, even though the transcript level of *CURCIN2* has decreased from the WE stage. In the endosperm, at 3 DAI a detectable amount of CURCIN2 can be observed, as a sequence of *CURCIN2* has been transcribed from 2 DAI. The continuous expression of *CURCIN2* until the WE stage results in the corresponding presence of CURCIN2 protein in the endosperm.

Band B of a protein has a slightly higher molecular weight than band C, which is CURCIN1. It is present in the embryo, seedling, stem, root and possibly cotyledon (very faint band). Because CURCIN3 and CURCIN4 cannot be recognized by the anti-curcin antibody (Figure 5.6), this protein is neither of them. It might be an unknown curcin protein or an unspecific protein, given that other curcin might exist in *J. curcas* (Table 5.4). There is also the possibility that band B could be the precursor of the mature CURCIN2 after the cleavage of the signal peptide but without any posttranslational modification. Wei (2005) has suggested that CURCIN2 is a glycosylated protein. Nonetheless, further studies such as immunoprecipitation are needed in order to elucidate which protein this band represents.

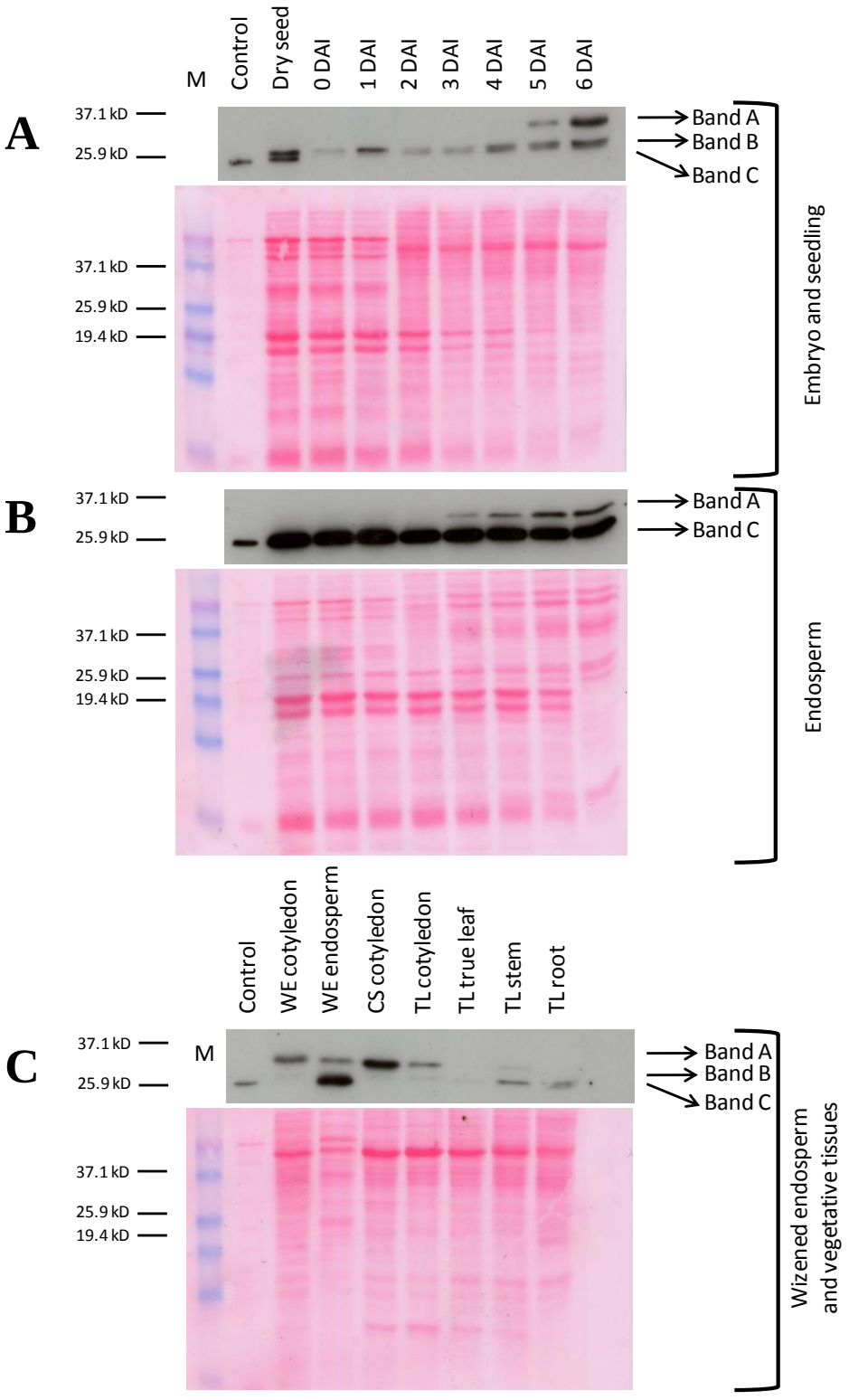


Figure 5.10 Immunoblot analysis and Ponceau staining of the possible curcin proteins in *J. curcas* during seed germination and seedling establishment (A and B), and in the vegetative tissues (C). Twenty micrograms of the total soluble protein extracts in each sample was loaded and anti-curcin was used as the primary antibody. Regarding the developing stages refer to Chapter 2. Ponceau stainings show equal loading. M: marker; C: protein extract containing CURCIN1 as positive control. DAI: days after imbibition. WE: wizened endosperm stage; CS: cotyledon spread stage; TL: true leaf stage.

5.5 Curcin2 is wound induced, and regulated by jasmonate

Previous studies have suggested that curcin2 is induced in a few stress conditions, including fungal infection, temperature (high and low), PEG-6000 (osmotic stress), UV light, ABA and SA (Wei et al., 2005, Qin et al., 2009b). In order to explore further how curcin accumulates upon response to wounding and salinity, we measured the curcin gene expression by qRT-PCR and protein abundance by immunoblotting. PEG-6000 treatment was included to enable comparison to the previous studies. Exogenous MeJA was applied to plants to discover whether JA would induce curcin expression. MeJA is volatile and this provides a simple way of supplying exogenous JA (Farmer and Ryan, 1990). Exogenous application of MeJA has been used in many studies to investigate the JA-responsive genes, especially when JA signalling deficient plants are not available (Suza et al., 2010).

To build a link between curcin and JA regulation, the expression levels of two putative JA synthesis genes in *J. curcas*, homologous to *LOX2* and *OPR3* in *Arabidopsis* were determined by qRT-PCR (these will be referred to as *LOX2* and *OPR3* in this thesis). Endogenous JA levels were measured in all stress treated plants except the MeJA treated plants, because of the possible carryover of the externally sprayed MeJA.

5.5.1 Stress induced curcin expression in J. curcas

As described in Chapter 2, 3-4 week old seedlings were subjected to the following treatments: mechanical wounding, NaCl, PEG-6000 and MeJA. I focused on *CURCIN2* since it is the major curcin gene expressed in the leaf material (Figure 5.9D). All treatments increased the expression of *CURCIN2* compared to the control samples (Figure 5.11). Wounding and NaCl treatments cause an almost 3-fold increase in *CURCIN2* transcript level as early as 15 mins from the onset of the treatment; a 7-fold increase was observed after 12 hours in the wounded leaves, and after 6 hours in NaCl-treated leaves (Figure 5.11A and B). Transcript abundance subsequently decreased in both cases. The transcript level of PEG-treated plants increases from 6 hours after treatment (Figure 5.11C). The maximum expression levels are reached at 12 hours (by 3-fold increase) and then the transcript abundance

decreases to base levels. In the MeJA-treated plants *CURCIN2* transcript level increases from 3 hours and peaks at 12 hours, at which time-point the transcript level is approximately 5-fold higher than the control sample (Figure 5.11D).

Since *CURCIN2* transcript level peaks at 12 hours for each of the treatments, the expression of *CURCIN1*, *CURCIN3* and *CURCIN4* were also assessed at this time point (and 15 mins as control). *CURCIN1* and *CURCIN3* transcript could not be detected in the control samples, nor under any of the stress treatments. The expression of *CURCIN4* is induced by about 8-fold under the wounding treatment at 12 hours, but not under any other treatments (Figure 5.12). Similar to the result of the time-course analysis (in Figure 5.9D), the expression level of *CURCIN4* is much lower than *CURCIN2* by about 100-fold.

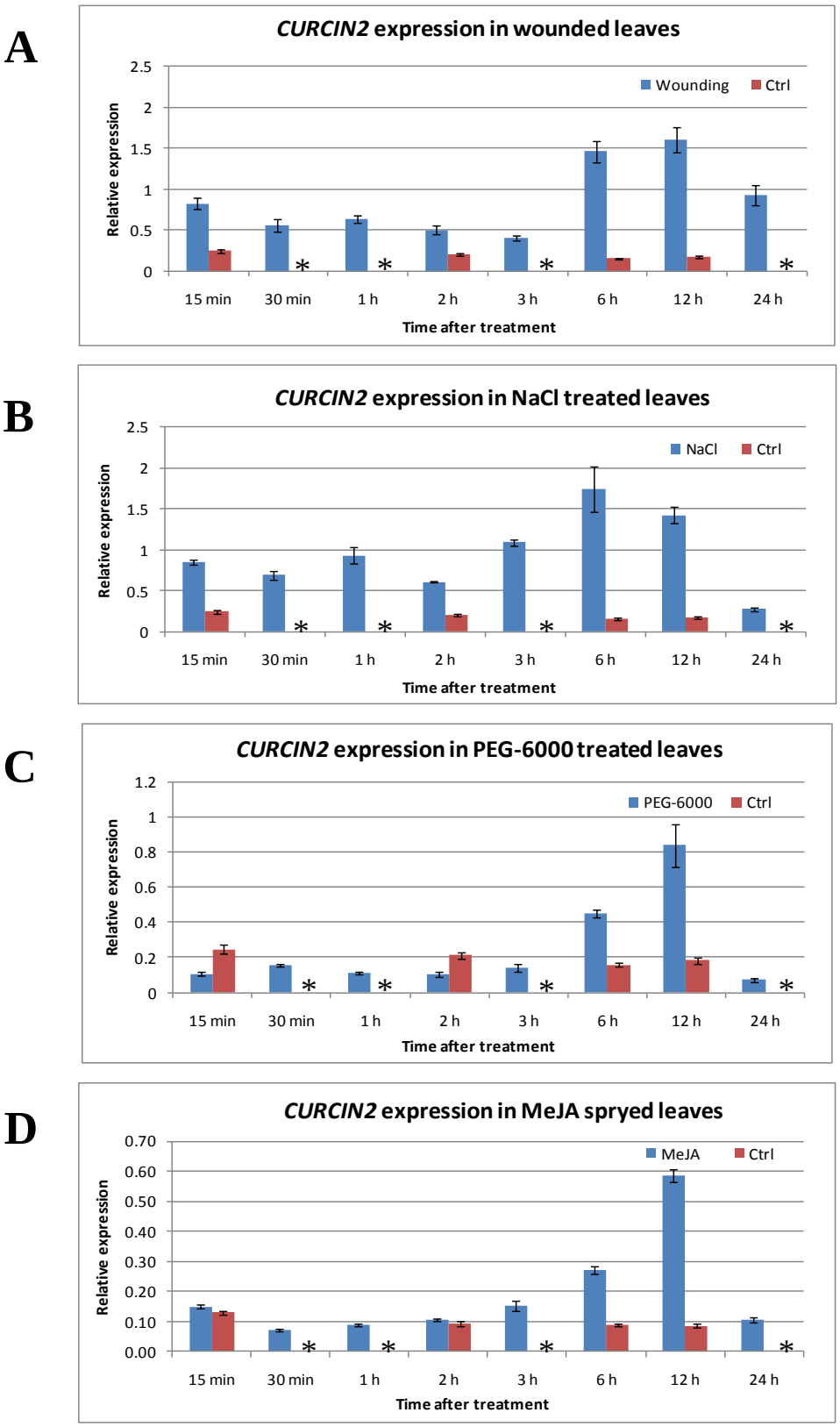


Figure 5.11 *CURCIN2* expression under stress treatments including wounding (A), NaCl (B), PEG (C) and MeJA (D) by qRT-PCR. All the expression levels were normalized against *ACTIN*. The error bars represent standard error mean (SEM) of 3 technical replicates. Ctrl: control. Asterisks: data not determined for control samples at 30 min, 1 h, 3 h and 24h.

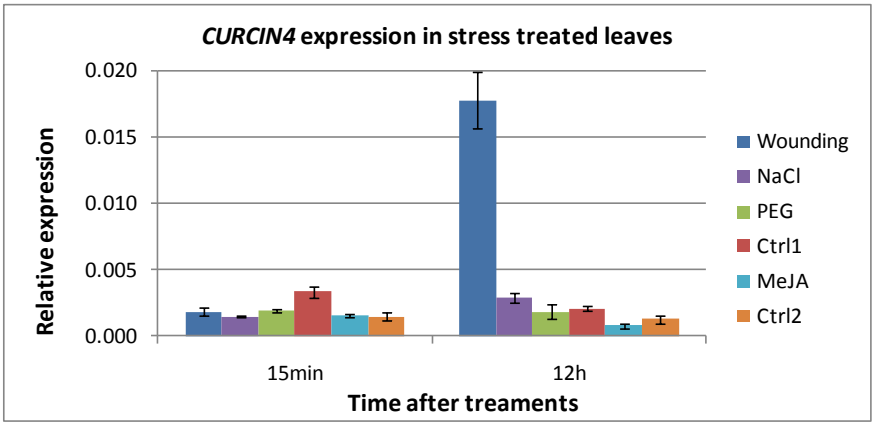


Figure 5.12 *CURCIN4* expression under stress treatments including wounding, NaCl, PEG and MeJA by qRT-PCR. All geneexpression levels were normalized against *ACTIN*. The **e**rror bars represent standard error mean (SEM) of 3 technical replicates. Ctrl: control.

5.5.2 Protein abundance of curcin under stress conditions

To determine the effect of stress treatments on the curcin protein level, protein extracts were prepared from the same batch of samples analyzed for qRT-PCR transcript level (Section 5.5.1). Proteins were electrophoresed on SDS-PAGE and subjected to immunoblot analysis with the anti-curcin antiserum. Two bands were identified by the anti-curcin antibody: band A and band B (predefined in Figure 5.10). As is shown in Figure 5.13, band A is absent or can hardly be detected in all the four untreated samples. However it could be identified after all the stress treatments. Together with the previous studies (Wei et al., 2005, Qin et al., 2009b), it is highly likely that band A represents CURCIN2. After wounding treatment (Figure 5.13A), CURCIN2 is induced after 15 min, and its abundance level peaks at 2 h and 3 h. With the exception of 1 h after treatment, CURCIN2 is induced in all the other stages analyzed. NaCl treatment induced the expression of CURCIN2 in all the stages from 15 mins to 24 hours, and a strong induction was found 6 hours after treatment (Figure 5.13B). PEG treatment induced the expression of CURCIN2 but only weakly (Figure 5.13C). The induction of CURCIN2 after the spraying of MeJA could only be observed 24 h after treatment (Figure 5.13D). The protein levels of CURCIN2 after the stress treatments exhibit some but not strict response to the transcript levels (Figure 5.11), suggesting that some posttranscriptional regulation or modification might have occurred.

Band B was detected in most of the samples, including treated and untreated samples. It is interesting that after wounding treatment, this band could not be detected in any of the treated samples. Under the other treatments, in some of the samples it was absent or found in a very low abundance. As mentioned in Section 5.4.2, it is not known if this band represents an unknown curcin protein, or an unspecific protein.

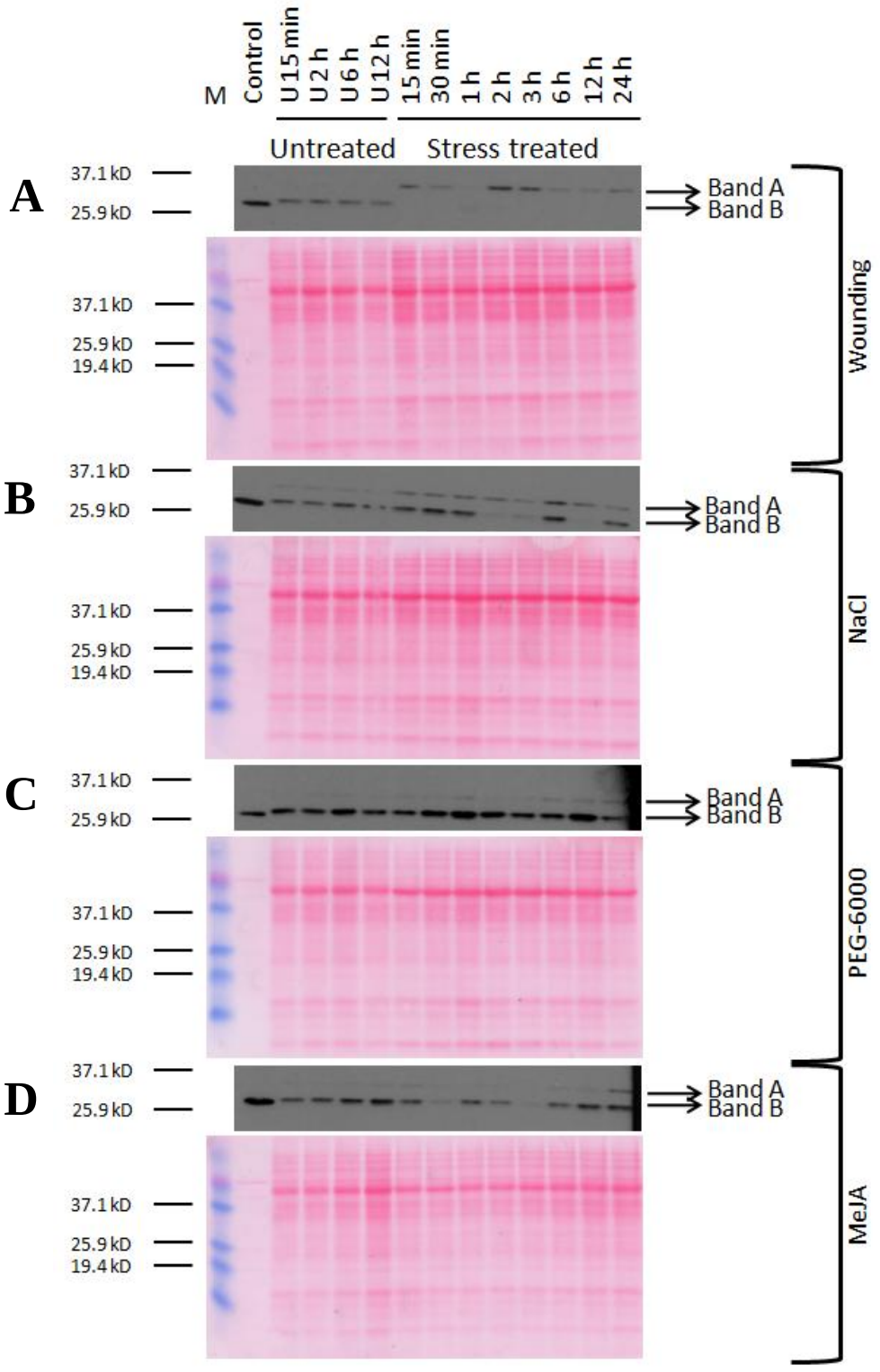


Figure 5.13 Immunoblot analysis and Ponceau staining of the protein abundance of curcin in *J. curcas* leaves under stress treatments including wounding (A), NaCl (B), PEG (C) and MeJA (D). Twenty micrograms of the total soluble protein extracts in each sample was loaded. The anti-curcin antibody was used as the primary antibody. M: marker; C: protein extract containing CURCIN1 as the positive control. Ponceau staining of the membranes shows approximately equal loading.

5.5.3 The expression of JA synthesis genes and JA accumulation and under stress conditions

In *J. curcas* two JA biosynthesis genes, homologous to *LOX2* and *OPR3* in *Arabidopsis* (will be referred to as *LOX2* and *OPR3* in this thesis) were identified and cloned by searching the 454 database previously made available in our lab. In order to investigate the possible regulatory role that JA might play in the induction of curcin under the stress treatments, the transcript abundance of *LOX2* and *OPR3*, and JA levels were determined in the plants after wounding, NaCl and PEG treatments. The transcript levels were also measured in plants subjected to exogenously applied MeJA.

After wounding, the expression of *LOX2* increases dramatically at 1 hour, increases steadily from 3 hours onwards and remains high for up to 12 hours (Figure 5.14A). The maximum expression of *LOX2* occurs at 6 hours after wounding by 100-fold compared to the control. An initial decrease in the *LOX2* transcript level was observed in the NaCl and PEG treatment (Figure 5.14B). Afterwards the *LOX2* transcript level goes back to base level in plants treated with PEG. In contrast, the NaCl treatment produces a 2-fold increase at 12 h. After the MeJA treatment increases in *LOX2* gene expression were observed at 2 h, 6 h and 12 h (Figure 5.14B). The maximum increase by *circa* 10-fold occurs at 12 h. The expression level of *OPR3* was found to be upregulated by wounding and MeJA but not by NaCl and PEG treatments (Figure 5.14C). The expression of *OPR3* in wounding- and MeJA-treated plants peaks at three hours after the onset of the treatments.

Wounding, but not any other stress treatments, elevates the endogenous JA level in the plants (Figure 5.14D). The JA level in the wounded plants increases from 1 h and peak at 2 h after the treatment by about 7-fold compared to the control. From 6 h after the treatment the level decreases gradually and remains constant. The levels of OPDA, the precursor of JA, were also examined in various stress conditions in this experiment, however no significant changes were observed.

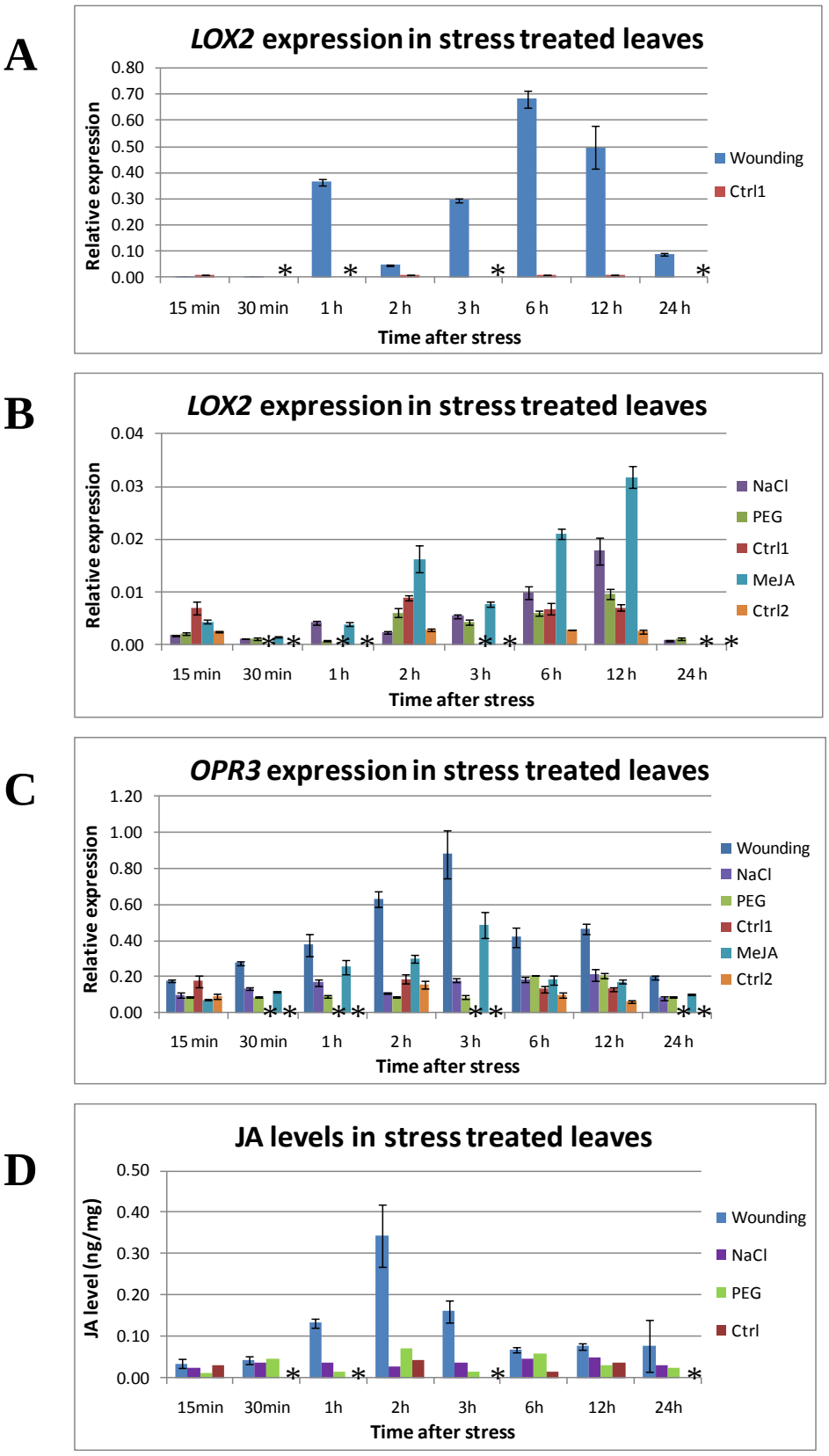


Figure 5.14 The transcript abundance of *LOX2* (A and B) and *OPR3* (C) and JA levels (D) in stress treated *J. curcas* leaves. In A, B and C, the error bars represent the standard error mean of 3 technical replicates. In D, the error bars represent the standard deviation. Asterisks: data not determined for control samples at 30 min, 1 h, 3 h and 24h. Stress treatments are as defined on each figure.

A summary is given regarding the effect of the different stress treatments on the transcript level of JA synthesis genes, curcin genes, and on endogenous JA levels in *J. curcas* leaves (Table 5.5). Wounding, NaCl, PEG and MeJA treatments all induce *CURCIN2* gene expression level, which also results in the correspondent induction at the protein level. Wounding also induces *CURCIN4* gene expression. The transcript abundance of two JA synthesis genes, *LOX2* and *OPR3* are elevated by wounding and exogenously applied MeJA, and the endogenous JA levels are increased under wounding. *LOX2* and *OPR3* expression, and endogenous JA levels do not increase after the NaCl and PEG treatments.

Table 5.5 A summary of the effect of the different stress treatments on the transcript level of JA synthesis genes, curcin genes, and on endogenous JA levels in *J. curcas* leaves.

	Wounding	NaCl	PEG	MeJA
LOX2	+	-	-	+
OPR3	+	-	-	+
JA levels	+	-	-	N/A
CURCIN1	-	-	-	-
CURCIN2	+	+	+	+
CURCIN3	-	-	-	-
CURCIN4	+	-	-	-

5.6 Discussion

5.6.1 Characterization of curcin in J. curcas

In this study, four curcin genes were cloned from the edible and non-edible *J. curcas* varieties. Two genes (*CURCIN1* and *CURCIN2*) have been previously reported (Lin et al., 2003a, Wei et al., 2005), whereas *CURCIN3* and *CURCIN4* are novel *J. curcas* genes. *CURCIN3* could only be amplified from the edible variety, whereas *CURCIN4* could only be amplified from the non-edible variety. A SNP was identified in *CURCIN1* of the edible and non-edible varieties. Although corresponding to the change of an amino acid (Met→Thr), this SNP lies in the coding region of the signal peptide, and is therefore unlikely to have an impact on the mature protein. There are no apparent differences in the MW of the mature CURCIN1 protein obtained from seeds of the two varieties. This demonstrates that

the amino-acid substitution does not affect processing of the protein. An intron was found in the 5'-UTR in each of the four *CURCIN* genes. This contrasts to the previous claims that curcin genes do not contain introns (Lin et al., 2003a, Wei et al., 2005). Since the intron is not in the coding region, it is likely that it was overlooked in the previous studies.

The exact number of curcin genes in this family is unknown. Zhang et al (2005) suggested that based on Southern blotting at least three copies of curcin genes exist in *J. curcas*. Yin et al (2010) cloned three curcin genes, of which two are identical to the *CURCIN1* and *CURCIN2* in this study. In the recently published genome sequencing of *J. curcas* by Sato et al (2011), two full-length curcin sequences are in agreement with the *CURCIN1* and *CURCIN2* in this study, and a few partial sequences (contigs) are contained in the full-length *CURCIN3* and *CURCIN4*. In addition, two curcin-like sequences are deposited in GenBank. The above studies suggest that there might be more curcin genes in *J. curcas*. Alleles are likely to exist in samples from different varieties. For instance, *CURCIN3* was only amplified from the edible Mexican variety, and *CURCIN4* was only amplified from the non-edible variety from Tanzania.

I further explored the spatial and temporal expression of the four curcin genes characterized in this study. The results show that *CURCIN1* is endospermic; *CURCIN4* is embryonic and expressed in the seedlings and the vegetative tissues; *CURCIN2* is more generally expressed in the endosperm, embryo and the vegetative tissues except root, and the expression level of *CURCIN2* is much higher than *CURCIN1* and *CURCIN4*; consistent with the prior PCR result (Section 5.2.1), *CURCIN3* expression could not be detected in the non-edible Tanzanian samples used in the time-course analysis. Compared to the recent published studies, Qin et al (2010) explored the promoter region of both *CURCIN1* and *CURCIN2*, and indicated that *CURCIN1* is endospermic, and *CURCIN2* is only inducible under stress treatment. However the expression of *CURCIN2* was detected in young *J. curcas* leaves without stress treatment by Yin et al (2010). The endospermic specificity of *CURCIN1* can be herein confirmed, and it is suggested that *CURCIN2* is not solely expressed upon stress. Since CURCIN1 is the main curcin in seeds, removal of it

would possibly reduce the seed toxicity in *J. curcas* without affecting other curcins in the vegetative tissues.

At the protein level, in spite of the low transcript of *CURCIN1*, CURCIN1 protein is abundant in the endosperm through germination and seedling establishment. This can be attributed to accumulation of CURCIN1 during the seed development stages, which has been substantiated by Qin et al (2010) and Yin et al (2010). To date, CURCIN1 is the only curcin isoform identified in the dry seed in *J. curcas*. However, since *CURCIN3* was cloned from the dry seed in the edible Mexican sample, it is not known if CURCIN3 accumulates to a large amount in the dry seed of the edible variety. CURCIN2 starts to accumulate in both the endosperm and seedlings a few days after seedling establishment until the wizened endosperm stage. Based on the fact that the transcript of *CURCIN4* was detected in the seedlings and the vegetative tissues, it is very likely that CURCIN4 is present in these tissues in *J. curcas*. Unfortunately CURCIN4 could not be detected by the anti-curcin antibody. The tissue specific distribution of curcin isoforms is reminiscent of the observation from other RIPs, including ricin from castor (Tully and Beevers, 1976), PAP from pokeweed (Poyet and Hoeveler, 1997), and maize b32 protein (Soave and Salamini, 1984).

The anti-curcin antibody identified another protein (band B) apart from CURCIN1 and CURCIN2. Three curcin gene sequences different from the four curcin genes in this study have been identified [from GenBank: FJ357424.1, JF357910.1; from Yin et al (2010): Sequence 3. JF357910.1 and Sequence 3 encode proteins with only 2 different AA.], and their AA sequences in the epitope region are identical to CURCIN2. Therefore, if these genes are not pseudogenes and encode curcin proteins, the unknown protein identified by the immunoblot analyses might represent one of these three curcins. Further studies are also needed to understand the numbers of curcin genes (and proteins) that exist in *J. curcas*.

Although the precise function of RIPs in plants is still the subject of debate (Nielsen and Boston, 2001), it is believed that they are related to plant defence. In addition, RIPs accumulate in plant tissues undergoing senescence, suggesting that they might have a function in regulating tissue senescence by arresting the metabolism

(Reinbothe et al., 1994a, Stirpe et al., 1996). This resembles the local suicide mechanism in the defensive role of RIPs, in which RIPs exert translation inhibitory activity against their own ribosomes. In this study, the abundance of CURCIN1 in the dry seed from the edible and non-edible varieties was examined. The lack of variation in the CURCIN1 content in the dry seeds of these two varieties of *J. curcas* confirms that the variation in seed toxicity can mainly be attributed to the presence of phorbol esters. CURCIN1 is present in the tegmen and endosperm tissues in *J. curcas* seed, implying that CURCIN1 may play a role in protecting the storage reserves in the endosperm from herbivores or attack from soil-borne microbes in the wild. It is conceivable that some roasting process must have been involved in the indigenous consumption of the edible *J. curcas* seeds in Mexico to inactivate CURCIN1. CURCIN2 accumulates from *de novo* synthesis during seedling establishment. Its biological function is unknown.

Each of the four curcin proteins contains a predicted signal peptide which is presumably involved in the subcellular localization of curcin proteins. In castor (*Ricinus communis*), the signal peptide directs the precursor of ricin into the ER (endoplasmic reticulum) lumen, to aid eventual mature ricin protein storage in the vacuole (Hartley and Lord, 2004a). Yin et al (2010) suggested that CURCIN1 is stored in the protein bodies but not in the apoplast, and CURCIN2 is possibly present in the vacuoles of leaf mesophyll cells and secondary cell walls of leaf immature tracheary elements, where the cells undergo programmed cell death. However, because of an unspecific substrate binding to the antibody used in their analysis, it is not clear whether CURCIN2 is present in both of these subcellular locations or only in one of the subcellular locations mentioned above.

5.6.2 JA accumulation in J. curcas upon wounding, and its regulatory role in the induction of curcin2 in wounding response

Studies in model plants, such as tomato and *Arabidopsis*, have elucidated JA biosynthesis, its signalling and its biological roles in plant defence response (Wasternack et al., 2006, Acosta and Farmer, 2010, Chini et al., 2009). In this analysis it was demonstrated that the expression of *LOX2* and *OPR3*, and JA levels are all increased after wounding and MeJA treatments. The increase of JA levels

occurs immediately after wounding, and peaks at 2 hours. This pattern is remarkably consistent with the previous study in *Arabidopsis* (Glauser et al., 2008, Reymond et al., 2000). The gene expression of *OPR3* shows a similar pattern to the change of JA levels, but with a 1 hour delay. Many studies have suggested that the JA synthesis genes are upregulated in response to wounding, but a burst in JA levels precedes this activation (Howe et al., 2000, Stenzel et al., 2003a, Stenzel et al., 2003b, Wasternack et al., 2006). Once subjected to wounding, plants use the pre-existing enzymes to initiate the JA synthesis, and because all genes encoding enzymes of JA biosynthesis are JA-inducible, a steady increase of JA levels will follow in later stages. In this study, the upregulation of JA synthesis genes, *LOX2* and *OPR3* by MeJA was observed in *J. curcas*.

The current study has shown that mechanical wounding, salinity, osmotic stress, and MeJA enhanced the expression of curcin2 at both the transcript level and protein level. This is in agreement with the recent published study by Qin et al (2010). Plants are very often subjected to wounding, salinity and osmotic stress, and the natural selection process must have endowed plants with various defences to protect themselves against those threats. Wounding is one of the common stresses to plants, caused by the herbivores and insects, which then leaves the plant susceptible to attack from microbes. As previously described, RIPs are plant defence proteins. Wounding induces JIP60 in barley (Reinbothe et al., 1994b), PIP2 in pokeweed (Song et al., 2000), saporins in soapwort (Tartarini et al., 2010), and in bitter melon (Xu et al., 2007). In *J. curcas* wounding induces curcin2.

Extensive studies have focused on the JA signal pathway in plant defence response and substantiated that the JA signalling pathway can regulate the wound response in plants (Cheong and Choi, 2003, Wasternack et al., 2006, Wasternack, 2007, Acosta and Farmer, 2010). Treatment of plants with JA enhanced their resistance to herbivore challenge (Howe and Jander, 2008), and also mutants deficient in JA biosynthesis or signal perception were more susceptible to herbivores (Wasternack, 2007, Zarate et al., 2007). In *J. curcas*, wounding induced JA biosynthesis as well as curcin2, and externally applied MeJA produced a similar effect. Together with the previous studies shown above, it can be concluded that it is very likely that wounding induces curcin2 via the JA signal pathway in *J. curcas*.

The NaCl and PEG treatments exhibited no influence on JA synthesis, although they induced curcin2 to a higher level. Apart from JA, a previous study suggests that OPDA may also be involved in wounding response in tomato (Farmer and Ryan, 1992). However no significant change of JA or OPDA was observed in the NaCl and PEG treatments. Therefore neither JA nor OPDA is involved in response to these stress conditions in *J. curcas*. The transcript level of *CURCIN4* is enhanced by the wounding treatment but not by exogenous MeJA. The above observation suggests that the induction of curcin2 at both the transcript and protein level under salinity and osmotic stress, and gene expression of *CURCIN4* under wounding, is regulated via other signal pathways (Bari and Jones, 2009).

5.6.3 Conclusion

In this study, four curcin genes were cloned and characterized. In the dry seed of *J. curcas*, CURCIN1 protein is of equal abundance in both the non-edible and edible varieties. This confirms that curcin is not the main toxic agent in the seed. Future attempts to detoxify *J. curcas* seed by biological means for the development of animal feed should focus on phorbol esters, but removal of seed curcin will also be beneficial. The tissue specific expression of curcin genes was demonstrated, suggesting that removal of curcin1 would not affect the plant defence equipped by other curcins in the vegetative tissues. Furthermore, *CURCIN2* is induced in *J. curcas* leaves under wounding, salinity and osmotic stresses. The JA signalling pathway functions when the plant is subjected to wounding, and based on this correlation it is concluded that JA is involved in the up-regulation of curcin2. Externally applied MeJA enhanced the expression of curcin2, further supporting this conclusion. In addition, this experiment raises the possibility of using MeJA or related compounds as sprays to enhance the resistance of *J. curcas* to biotic or abiotic stresses mentioned above.

Chapter 6: Discussion

In this study, biochemical analyses of *J. curcas* seed composition were performed. Phorbol esters are absent in the edible *J. curcas* variety, but seed curcin (CURCIN1) levels are similar in both varieties, which confirms that seed toxicity is likely due to the single phorbol ester trait. This should simplify the detoxification process. The AFLP result shows that the edible and non-edible varieties are genetically different. Furthermore, the edible samples group into a single cluster, which suggests that these samples share a common mutation, perhaps in the phorbol ester biosynthesis pathway. The AFLP result also suggests that future germplasm selection should be focused on the Meso-America regions which have more genetic variation.

A significant variation in seed mass, seed oil content and seed oil composition in Madagascan samples, which appear to be almost identical by AFLP analysis, could be due to either environmental or, as recently reported, epigenetic effects (Yi et al., 2010). No significant correlation was observed between either seed mass or oil content and available environmental data. This suggests that more field trials are needed to explore how environmental conditions affect key traits. With more *J. curcas* plantation being expected in the next a few years, field data should help answer the above question. However, temperature was found to play an important role in determining the oleate and linoleate content. This result shows the variation in the oleate and linoleate content observed by Makkar and Becker (2009) is also likely owing to environmental variability rather than genetics claimed by the authors. High oleate content is known to result in higher cetane values, which is desirable for biodiesel (Knothe, 2008). Selecting sites with the appropriate growth temperature may be necessary to produce *J. curcas* oil with desirable fatty acid composition if provenances with a genetic composition similar to those grown in Madagascar are used.

In this study, only a small amount of seeds (<10) were available from the edible provenances, and therefore it was impossible to characterize the oil content and fatty acid composition. A study by Makkar and Becker (2009) suggested that there is no apparent difference between the edible and non-edible varieties regarding oil quality and quantity. However, as environment factors contribute to both of these traits, the

Mexican edible variety should be tested alongside other varieties to properly evaluate performance.

Despite the global interest in *J. curcas* as an oilseed crop over the last two decades, work on developing new varieties of the plant has commenced only very recently [D1 Oils plc (http://www.d1plc.com/agronomyBreeding.php) and Yang et al (2010)]. However, a number of traits have been identified in *J. curcas* which could be improved (Divakara et al., 2010). As well as seed edibility, which has already been discussed, seed oil yield is perhaps the most economically important trait. In *J. curcas*, seed oil yield improvement could be achieved through increasing female/male flower ratio, number of branches, seed size or oil content.

Conventional breeding approaches, such as mass selection, recurrent selection and hybridization have proven successful in many plant species (Acquaah, 2007). Using mass selection, a small scale field trial was initiated in 2005 to select high seed yield and oil varieties from 80 provenances from China, and a provenance with higher yield has been claimed in this study (Yang et al., 2010). However it is not clear if this provenance is genetically distinct, because no prior genetic diversity study was conducted, and no statistical support was provided.

There are few reports on the creation of hybrid *J. curcas*. Sujatha et al (2005) have shown that crossing a maternal edible variety with a paternal non-edible variety gives rise to edible seeds, but using a non-edible variety as maternal parent produced non-edible seeds. This implies that phorbol ester synthesis in *J. curcas* might be maternally controlled. Given the high genetic variation within the Jatropha genus (Basha and Sujatha, 2009), inter-specific hybridization could provide a feasible way to introduce new alleles into *J. curcas*. Inter-specific hybrids resulted in improved fruit set and increased fruit size (Basha and Sujatha, 2009), reduction in phorbol ester content, increased number of fruits per inflorescence and increased seed weight, oil content and oil composition (Popluechai et al., 2009). While further backcrosses are needed to confirm the genetic basis of these traits, the initial results look promising.

Given the narrow genetic diversity so far reported for *J. curcas*, and the long generation time, conventional breeding efforts could be both difficult and very lengthy. However, modern plant breeding techniques, which rely on genetics and biotechnology, can rapidly shorten the development time for new plant varieties. Such approaches, including mutation breeding, plant transformation and marker assisted selection, are also more precise compared to the conventional methods (Acquaah, 2007). Since the genetic diversity in *J. curcas* may be low, the generation and selection of plants containing desirable mutations can be achieved by the use of mutation breeding techniques such as TILLING (Targeting Induced Local Lesions in Genomes) (Colbert et al., 2001). This technique has been applied to many crop species, such as maize, rice, soybean and wheat (Cooper et al., 2008, Slade et al., 2005, Till et al., 2007, Till et al., 2004). In *J. curcas* possible targets could be to reduce toxicity and to alter the fatty acid composition by disrupting phorbol ester biosynthesis genes such as terpene cycleases, and fatty acid desaturase *(FAD2)*.

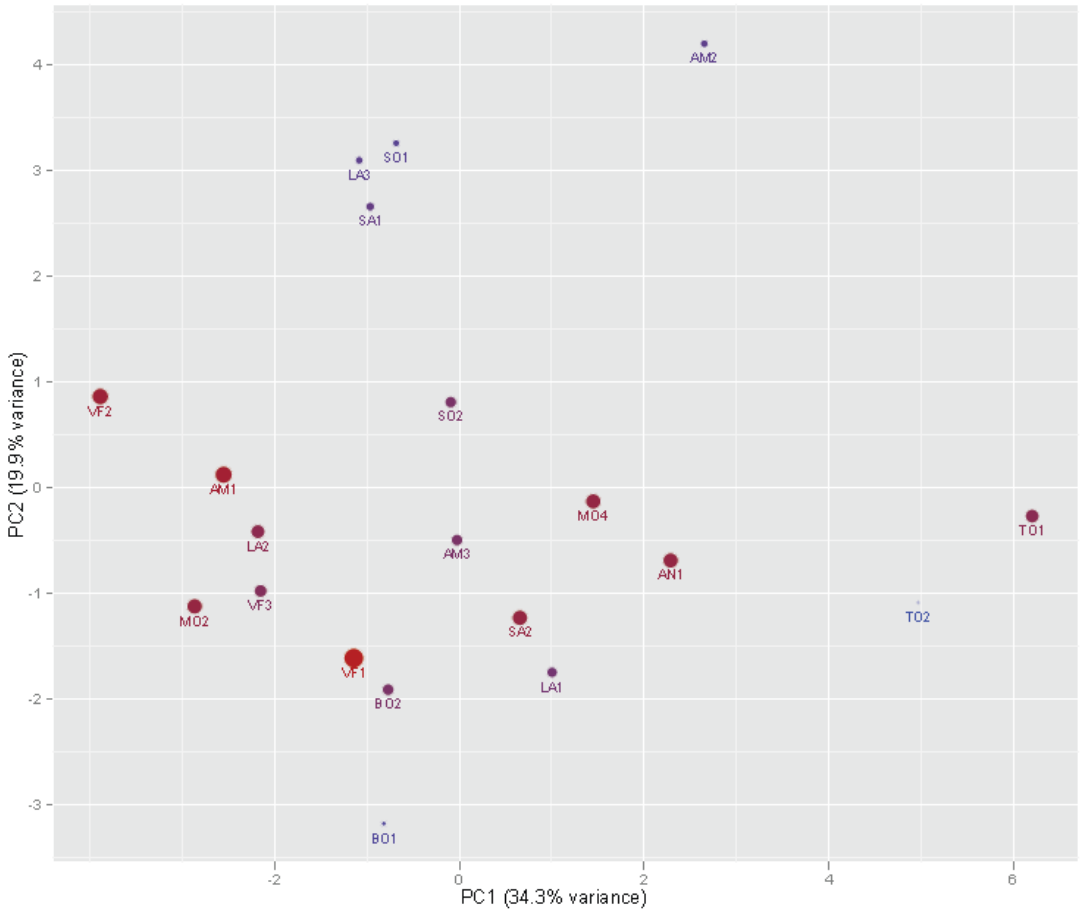
A large amount of DNA sequence information has become available recently from transcriptome sequencing projects (Costa et al., 2010, King et al., 2010, Natarajan et al., 2010, Natarajan and Parani, 2011, Sato et al., 2011). Furthermore, approximately 70% of the *J. curcas* genomic sequence is now available (Sato et al., 2011). These sequence databases will serve as a useful tool to allow the study of the key genes involved in processes such as oil synthesis, phorbol esters synthesis (see Section 3.4) and flowering control.

The AFLP markers produced in this study include several markers unique to the edible and non-edible varieties, and thus can be a useful resource in addition to SSR markers for the creation of a genetic linkage map of *J. curcas*, and the identification of markers associated with desirable traits. Such genetic maps are available for most crop species (Acquaah, 2007), and are often used to develop new improved plant varieties with desirable gene combinations.

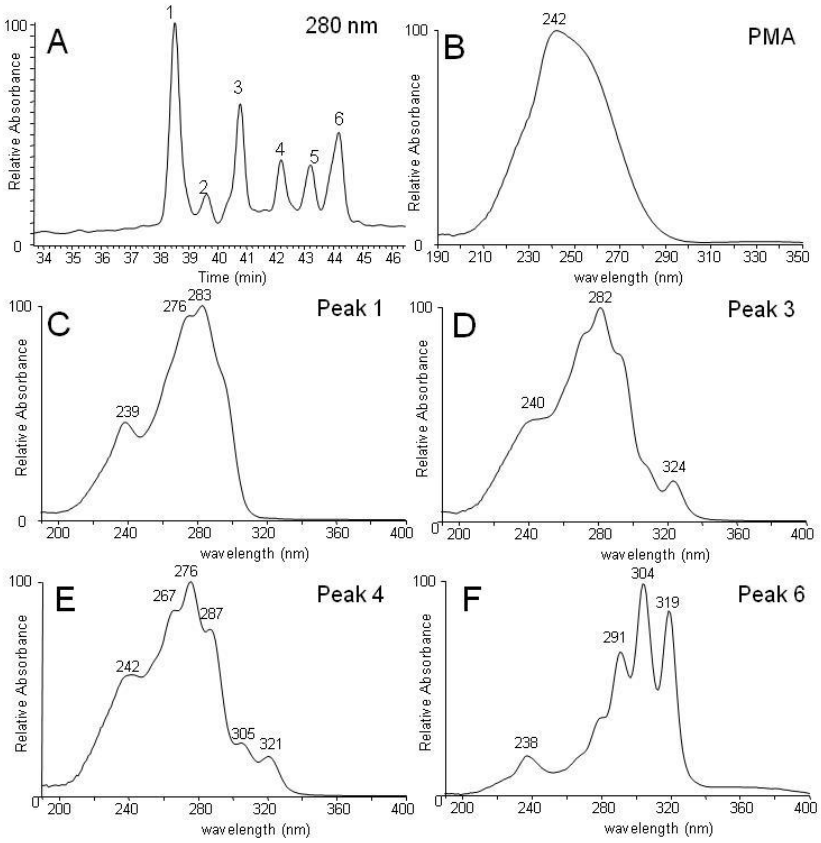
Given the progress in *J. curcas* research which has been made in this study and also in other groups, in the near future, constructing a genetic map to assist breeding and mining genes of key biological interests will be the two main directions for *J. curcas*

development. Both approaches are now well placed to underpin the development of *J. curcas* as a valuable industrial crop.

Appendices



Appendix 1 PCA analysis of the average oil content against the environmental data. Dimension of the dots is proportional to the average oil content data.



Appendix 2 A: HPLC chromatograph of *J. curcas* phorbol esters obtained at 280 nm. B: UV spectrum of the internal standard, phorbol-12-myristate-13-acetate (PMA). C-F: UV spectra of four of the *J. curcas* phorbol esters indicated on panel A. The λmax (nm) for each of the peaks is indicated on the trace. This figure was produced by Dr. Andrew King.

Appendix 3 Seed mass, kernel percentage and oil content of 23 *J. curcas* provenances from Madagascar.

Accessions	Seed mass					Kernel %			Oil Content				
	AVE	SD	SEM	95% CI-	95% CI+	AVE	SD	SEM	AVE	SD	SEM	95%CI-	95%CI+
AM1	634	136.2	9.6	614.7	652.5	64.7%	2.7%	0.5%	33.4%	3.9%	0.7%	32.0%	34.8%
AM2	670	126.4	8.9	652.1	687.1	63.2%	7.3%	1.3%	30.5%	3.2%	0.6%	29.4%	31.7%
AM3	722	127.0	9.0	704.5	739.7	64.0%	3.9%	0.7%	30.3%	3.8%	0.7%	28.9%	31.6%
AN1	659	120.4	8.5	642.2	675.6	64.4%	2.4%	0.4%	32.8%	2.6%	0.5%	31.8%	33.7%
BO1	558	137.3	9.7	538.5	576.6	54.0%	7.2%	1.3%	24.2%	4.8%	0.9%	22.5%	25.9%
BO2	612	136.1	9.6	592.7	630.4	57.7%	5.0%	0.9%	29.2%	5.0%	0.9%	27.4%	31.0%
LA1	672	136.2	9.6	653.6	691.3	64.8%	3.4%	0.6%	31.6%	2.3%	0.4%	30.7%	32.4%
LA2	578	141.3	10.0	558.0	597.2	63.1%	5.4%	1.0%	32.2%	4.2%	0.8%	30.7%	33.8%
LA3	624	146.1	10.3	603.3	643.8	65.3%	2.8%	0.5%	30.6%	2.9%	0.5%	29.6%	31.7%
MO1	657	122.6	8.7	640.2	674.2	64.3%	5.4%	1.0%	35.2%	4.2%	0.8%	33.7%	36.7%
MO2	688	122.7	8.7	671.2	705.2	64.3%	3.7%	0.7%	35.3%	2.4%	0.4%	34.4%	36.2%
MO3	619	148.2	10.5	598.3	639.4	63.5%	6.0%	1.1%	32.6%	3.9%	0.7%	31.2%	34.0%
MO4	634	110.3	7.8	619.0	649.6	65.6%	1.9%	0.3%	32.7%	2.4%	0.4%	31.8%	33.5%
MO5	568	143.4	10.1	548.4	588.2	62.1%	5.4%	1.0%	33.8%	4.0%	0.7%	32.4%	35.3%
SA1	643	148.8	10.5	622.4	663.7	60.3%	5.6%	1.0%	27.7%	6.6%	1.2%	25.3%	30.0%
SA2	558	137.3	9.7	538.5	576.6	62.0%	3.0%	0.5%	32.1%	3.2%	0.6%	31.0%	33.2%
SO1	681	105.8	7.5	666.1	695.5	66.3%	3.0%	0.5%	35.7%	2.5%	0.5%	34.8%	36.6%
SO2	546	143.9	10.2	525.6	565.5	65.2%	3.8%	0.7%	33.4%	3.7%	0.7%	32.1%	34.8%
TO1	653	78.6	5.6	642.1	663.9	66.1%	4.0%	0.7%	33.8%	2.6%	0.5%	32.9%	34.7%
TO2	649	149.9	10.6	628.1	669.6	58.2%	10.2%	1.9%	24.1%	6.1%	1.1%	21.9%	26.3%
VF1	619	148.2	10.5	598.3	639.4	61.9%	3.5%	0.6%	37.7%	3.7%	0.7%	36.4%	39.0%
VF2	596	150.7	10.7	575.6	617.4	65.8%	3.9%	0.7%	34.5%	4.0%	0.7%	33.1%	36.0%
VF3	743	107.8	7.6	727.7	757.5	62.5%	4.5%	0.8%	35.8%	3.9%	0.7%	34.3%	37.2%

Appendix 4 Genomic DNA sequence of the four curcin genes. Primers are shown in underlined italic (see Chapter 2). Red letters show the ORF region (does not include the stop codon). Black boxes show the introns.

>*CURCIN1*

ACCATCTCTTGCTCTCTTCTTCTTTACTTCCCCGT*TTGCTCAGTTGCTTTCTT*
*T*GTAAGTAATATTGAAGCCTCTGCCCTTCTTTTTTGTTGACAAATTCCATT
TTTTTGTTTTACTAATAGCATGTTAATTTCTAGCTTCTGGAAATGAGTTTA
TTATCCTTTATATGATAAACTTGTGACCATTCTATCTCTTTTTAATTATTTT
TATAATTTTATGCAATTCTATTAAAATAATCGTATTCGTATAATGATATTT
GTGTTTCTTCATACAACTGGACAGGTGAAATCAATATGAAAGGTGGAAA
GATGAATCTCTCCATTATGGTGGCTGCATGGTTTTGCTGGAGTAGTATTA
TATTCGGATGGGCATCGGCTAGGGAAATAGTTTGTCCATTCTCATCAAAC
CAAAACTACAAAGCTGGTTCCCCTCCAACTTTAACCATTACTTATGACGC
TACTACTGATAAGAAAAACTACGCCCAGTTCATTAAAGATCTAAGAGAA
GCATTTGGCTTCAGTTATTCAAGCCATGAAATACCAGTCTTACGGGCCAC
AGTTGCTCCAAATCAGAAATTTATTGTAGCCAAAGTCATAAATGTAGCGA
ATTTAGAAGTATCATTAGGATTAAACGTCGTTAATGCGTATTTAGTGGGT
TATAAGGTAGGAGGTACTTCCTATTTCTTTAACGATCCGGAATCTTTGGC
TGATGCAAAAACATATCTTTTCACAGACACAAAGCAACAAACGCTATCA
TTTACTGGTAGCTATGCAGATTTTCTATCTAGGGCAAACGTACACAGAGA
GGATGTGGATTTAGGGGTGCAGGCATTAGATAATTACATATATACACTTG
AAAAAAGTTCAAAGCCAGCAGACATTGCTAAACCTCTAGTTGGTTTTATC
GAAATGGTTCCAGAGGCAGCAAGATTCAAATATATTGAGAAAAAAGTAT
TAAGTCAAATTAGCAAAACCTTTAGGCCGGGTGGTGACATAATTAGCCTT
GAGAACAACTGGGGAGACCTCTCTTATCAAATACAGAAATGTGTAAATG
GTGTATTTCTGAAGCCAGTTCAATTACAACGTGAAAACTATACCAATATC
CTAGTGAACAATGTCACCCAAGTAGCAGGTGTCATGGGAGTCTTGTTGA
ATGCAGTCAATTACAAAGTCTGAATGGAAGAAATTATTTTCAACTACCAA
AAGTGGCTGCCATGGCTTTAATCCTACTTTTGCTCTATATATAGAGTAGC
ATAAATAAAGGACAACAAATTTATTATTATTGTTGCTAATGCTATATGCT
ATTTCCCTGTAATATCCTCATCTTTCCAATGTATGAATATGATGATGAATT
ATATATGA*CAAATAAAGTTTCTACTAGTTC*TTAATTACG

The SNP is shown in bold at 17 bp-C in Mexican samples, T in Tanzanian samples. ORF length: 879 bp

>*CURCIN2*

*TTGCTCAGTTGCTTTCTTTG*TAAGTAATATTGAAGCCTCTGCCCTTCTTTTTT
GTTGACAAATTCCATTTTTTTGTTTTACTAATAGCATGTTAATTTCTAGCT
TCTGGAAATGAGTTTATTATACTTTATATGATAAACTTGTGACCATTCTAT
CTCTTTTTAATCATTTTTATAATTTTATGCAAATCTATTATAATAATCGTA
TTCGTATAATGATATTTGTGTTTCTTCATACAACTGGACAGGTGAAATCA
ATATGAAAGGTGGAAAGATGAACCTCTCCATTATGGTGGCTGCCTGGTTT
TGCTGGAGTAGTATTATATTCGGATGGGCATCGGCTAGGGAAATAGTTTG
TCCATTCTCATCAAACCAAAACTACAAAGCTGGTTCCACTCCAACTTTAG
CCATTACTTATGACGCTACTACTGATAAGAAAAACTACGCCCAGTTCATT
GAAGATCTAAGAGAAGCATTTGACTTCAGTTATTTAAGCCATAAAATACC
AGTCTTACGGGCCACGGTTGCTGCAAATCAGAAATTTATTGTAGCCAAAG
TCATAAATTCTGGGGACATAGAAGTATCAGTAGGATTAAACGTCATTAAT
GCATATCTAGTGGCTTATAAGGTAGGAAGTAATTCCTATTTCTTTAACGA
TTCGGAATCTTTGGCTGATGCAAAAAAAAATCTTTTCACAGACACAAACC
AACAAACACTAGCATTTACTGGTAGCTATGCAGATTTTGAATCTAGGGCA
AAGTTACATAGAGAGGAAGTGGATTTAGGAGTGGTGGCATTGGATAATT
ACGTATATACACTTGAAAAAAGTTCTCAGCCAGCAGACATTGCTAAACCT
CTAGTTGGTTTTATCGAAATGGTTCCAGAGGCAGCAAGATTCAAATATAT
TGAGAAAAAAATATCAACTCAAATTAGCAAAACCTTTAGGCCGCGTGGT
GACATAATTAGCCTTGAGAACAACTGGGGAGACCTCTCTTATCAAATAC
AGAAATCTGTTGATGATGTATTTCTGAAGCCAGTTCAATTGCAACGTGAA
AACTATACCAATATCCTAGTGAACAATGTCACCCAAGTAAAAGGTCTCAT
GGGAGTCTTGTTGAATGCAGTCAATTACAAAGTCTCAATGGAAGAAATT
ATTTTCAACGACCAAAAGTGGCTGCCATGGCTTTAATCCTACTTTTGCTCT
ATATATAGTAGCATAAATAAAGGACAACAAATTTAGTATTATTGTTGTTG
TCCAAACATGTTGCTAATGATATATGCTCTTTCCCTGTAATATCCTCGTCT
TTCCAATGTATGAATATGATGATGAATTATATATGA*CAAATAAAGTTTCTAC*
TAGTTC

ORF length: 927 bp

>*CURCIN3*

*CATTAATTCCTTCTTGCTCTTGCG*CTCTTTGTAAGTTATCTTTTTGATAAATT
CCTCTAACATTAACTTCCCATTTCTTCTTTCACTAACAGCCTTTACCCAAA
ATGTTAAACACTACGATCATTCTTTCTTTTCCTAATTATTTTTATAATTTT
ATGCAATTTTATTAAATAACTATATTTATGTAATGATATCTGTGTTTCTTC
ATGCAAATGTTTTTATTTGTATGATAATTAATTAAGCATTTTGACAGGTG
AAATCAATATGAAACGAGGAAACACGAAGCTCTGCATTATGGTGGCTAC
ATGGTTTTGCTTGAGTACTATTATATTTGGATGGGTATCGGCTAGGGAAA
TAGATTCTCCATTCTCATCAAACGACCAGTACACAGCTGATTCCGTTCCA
ACTTTAACCATTACTTATGATGCTGATAAAGATAAGCAAAACTACGCCAA
GTTTATTAAAGATCTAAGACAAGCATTTGGCTCCAGCGGTTTAAGCCACG
GAATACCAGTCTTACGGGCCTCAGTGGCTGCAAATCAGAAATTTTTTGTA
GCCAAAGTCATAAATTCAGGGGATATAGAAGTATCAGTAGGATTAAACG
TCATTAATGCATATCTAGTGGCTTATAAGGTAGGAAGTAAATCCTATTTC
TTTAACGATACGAATTTTTTGGCCGATGCAAAAAAATATCTTTTCACAGA
CACACAGCAACAACCAGAATTATCATTTACTGGTAACTATGCAAATTTTG
AATCTAGGGCTAAGGTACGTAGAGAGCAAGTGGATTTAGGAGTGGTGGC
ATTAGATACTTACATATATGATCTTCAAAAAAGTTCTCTACCAGCAGACA
TTGCTAAACCTCTAGTTTCTTTTATCCAAATGGTTTCAGAGGCAGCCCGA
TTCAAATATATTGAGAACAAAGTATTAGCTCAAATTAGCCAAAAATTTAG
GCCGAGCGGTGACATACTTAGCCGGGAGAACAGTTGGGACGACCTATCT
TATCAAATACAGAAATCTGTAAATAATGTATTTCTCAACCCAGTTCAATT
GCAACGTGAAGACCATTCCTTTTACCAAGTGAACAATGTCAACCAAGTA
AAAGATGATATGGGAATCCTGTACAGTGAAGCCAATCACAAAGTCTCCA
CGGAAGAAATTATTATCAACAGCTAAAAGTGGCTGCCAATGCTTTAATA
GTACTATGCTGTAAAGATAGTAGCATCAATAAAGGACGTCATTGTCCAA
GCATTCTGCTAATGCTTTGCTATTTCCCTTTAATTTCCTCATATTTCCAAT
GAGAATAAAGACTTATAGTAGGTTATATAAATTATTGTCAGATTATGTCT
TGAATTTNGTGTGTGACTTTGCCTTCAAGCCCAATG*TTTGTATTGCTTTGCT*
TGGGTATAAAGCATGAATTAAAG

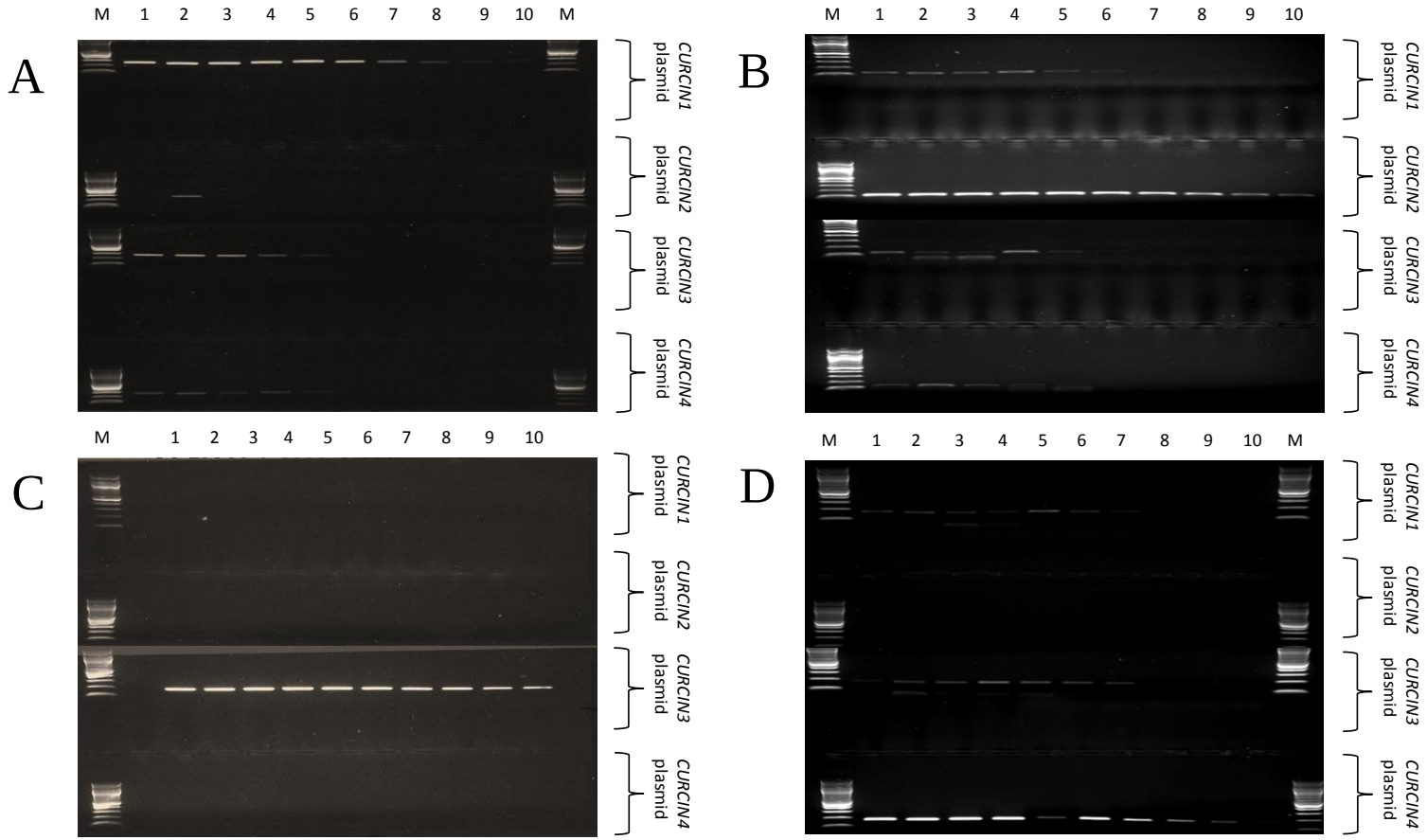
ORF length: 909 bp

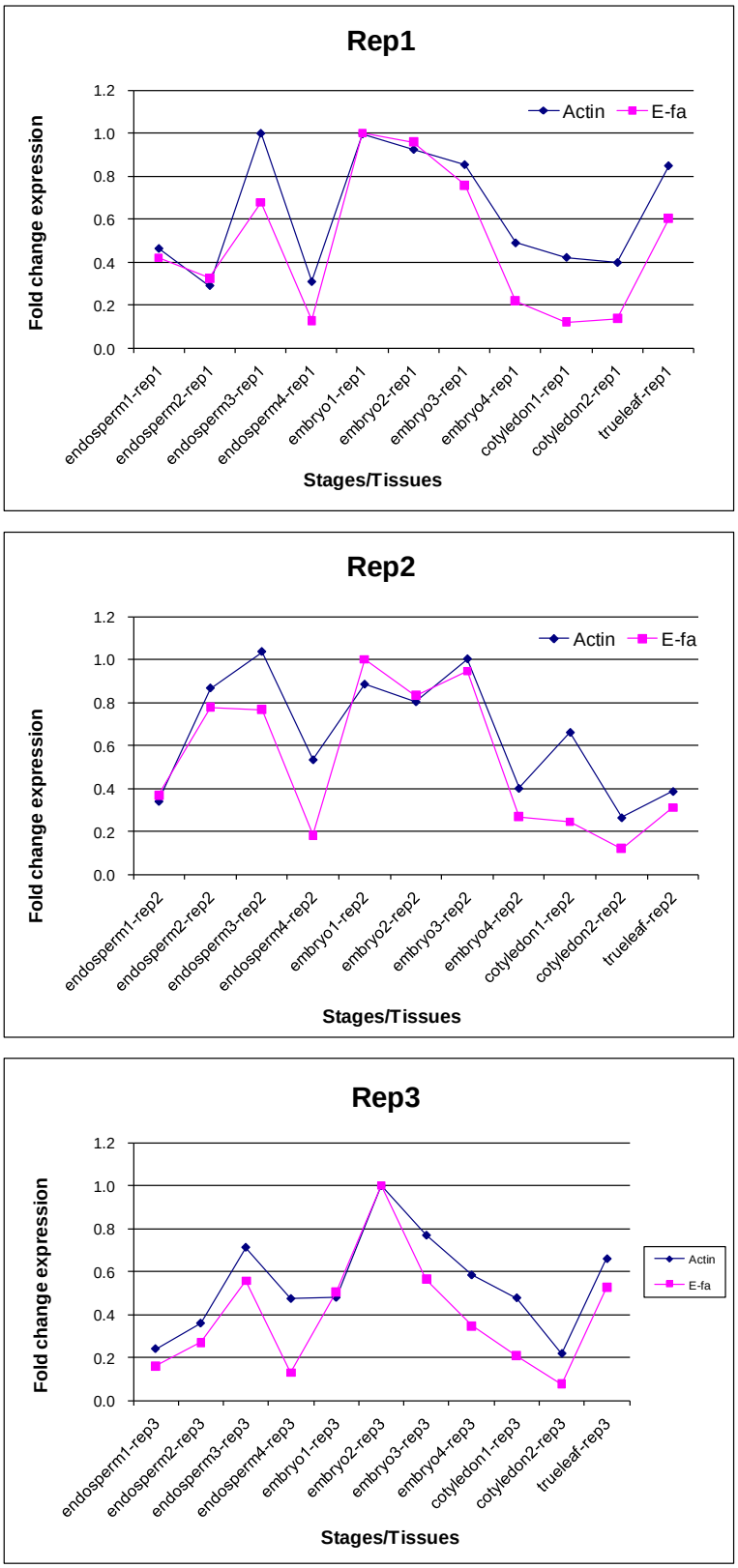
>*CURCIN4*

*CATTAATTCCTTCTTGCTCTTGCG*CTCTTTGTAAGTTATCTTTTTGATAAATT
CCTCTAACATTAACTTCCCATTTCTTCTTTCACTAACAGCCTTTACCCAAA
ATGTTAAAAACTACGATCATTCTTTCTTTTCTTAATTATTTTTATAATTTC
ATGCAATTTTATTAAATAACTATATTTATATAATGATATCTGTTTTTCTTC
ATGCAAATGTTTCTATTTCTATGATAATTAATTAAGCATTTGGACAGGTG
AAATCAATATGAAAGGTGGAAACATGAAGCTCTGCATTATGGTGGCTGC
ATGGTTTTGCTGGAGTACTATTATATTTGGATCGGCATCCGCTAGGGAAA
GAGCTTGGCCATTCTCATCAAACAACAACTACGCAGCTGATTCCACTCCA
ACTTTAACCATTACTTATATTCCTGATGAGGATGAGCAAAACTACGCCAA
GTTTATTACAGATCTAAGAGAAACATTTGGCTCTAGCGGTTTAAGCCATG
GAATACCAGTCTTACGGGCCACAGTTGCTGCAAATCAGAAATTTTTTGTA
GCCAAAGTCATAAATGCTGGGGATATAGAAGTATCAGTAGGATTAAACG
TCATTAATGCATATCTAGTGGCTTATAAGGTAGGAAGTAATTCCTATTTC
TTTAACGATTCGGAATCTTTGGCTGATGCAAAAAAATATCTTTTCACAGA
CACAAAGCAACAAACACTAGCATTTACTGGTAGCTATGCAGATTTTGAAT
CTAGGGCAAAGGTACATAGAGAGGAAGTGGATTTAGGAGTGGTGGCATT
AGATAATTACATATATGATCTTCAAAAAAGTTCTCTACCAGCAGACATTG
CTAAACCTCTAGTTTCTTTTATCCAAATGGTTTCAGAGGCGGCCCGATTC
AAATATATTGAGAACAAAGTATTAGATCAAATTAGCCAAACATTTAGGC
CGCGCGGTGACATACTTAGCCGGGAGAACAGCTGGGACGACCTATCTTA
TCAAATACAGAAATCTGTAAATGATGTATTTCTCAACCCAGTTCAATTGC
AACTTGAAGACTATTCCTTTTACCAAGTGAACAATGTCAACCAAGTAAAA
GATGATATGGGAATCCTGTACAATGAAGCCAATCACAAAGTCTCCATGG
AAGAAATTATTATCAACAGCCAAAAGTGGCTGCCATTGCTTTAATAGTAT
TACTTTTGCTCTATCTTTGCTGCTCTAAATATAGTAGCATCAATAAAGGA
CATCATTGTCCAAACATTCTACTAATGCTTTGTTATTTCCCTGTAATTTCC
TCATATTTCCAATAAGAATAAAGACTTATAGTAGGTTTTATAAATTATTG
TCGGATTATGCCTTGAATTTTGTGCGTGACTTTGCCTTCAAGCCCAATG*TT*
TGTATTGCTTTGCTTGGG

ORF length: 927 bp

Appendix 5 Gradient PCR to test the primers' specificity to *CURCIN1* (A), *CURCIN2* (B), *CURCIN3* (C) and *CURCIN4* (D). M: marker. 1-10: gradient PCR with different *T*a (54 ºC, 54.9 ºC, 55.7 ºC, 56.8 ºC, 58.3 ºC, 60 ºC, 61.4 ºC, 62.5 ºC, 63.3 ºC, 64 ºC).





Appendix 6 Relative gene expression (normalized against the highest value) of *ACTIN* and *ELONGATION FACTOR1α* during seed germination and seedling establishment in *J. curcas*. Rep: replicate.

Abbreviations

Å	Angstrom
ABA	Abscisic acid
AFLP	Amplified fragment length polymorphism
bp	Base pairs
cDNA	Complementary DNA
CN	Cetane number
CoA	Coenzyme A
CTAB	Cetyltrimethylammonium bromide
DAI	Days after imbibition
DNA	Deoxyribonucleic acid
dNTP	Dinucleotide triphosphate
DTT	Dithiothreitol
DW	Dry weight
EDTA	Ethylenediaminetetraacetic acid
EST	Expressed sequence tag
EU	European Union
EV	Empty vector
FA	Fatty acid
FAMEs	Fatty acid methyl esters
g	Grams
g	Gravity
GC	Gas chromatography
GHG	Green house gas
H_3PO_4	Phosphoric acid
His	Histidine
HCl	Hydrochloric acid
HPLC	High performance liquid chromatography
HRP	Horseradish peroxidase
IgG	Immunoglobulin G
JA	Jasmonic acid
ICL	Isocitrate lyase
JA-Ile	Jasmonoyl-L-isoleucine
IPTG	Isopropyl β-D-1-thiogalactopyranoside

ISSR	Inter-simple sequence repeat)
Kan	Kanamycin
kb	Kilo base pairs
KCl	Potassium chloride
kD	Kilo Dalton
kg	Kilogram
L	Litre
m	Metre
LC-MS	Liquid chromatography-mass spectrometry
LOX	Lipoxygenases
MeJA	Methyl jasmonate
LB	Lysogeny broth
M	Molar per liter
LD_{50}	Median lethal dose
m^3	Cubic meter
mg	Milligram
min	Minute(s)
MLS	Malate synthase
mRNA	Messenger RNA
MS	Mass spectrometry
µl	Microlitre
µm	Micrometre
µM	Micromolar
MALDI-TOF	Matrix assisted laser desorption/ionization-time of flight
MCMC	Markov Monto Carlo Chain
MI	Marker Index
ml	Millilitre
mm	Millimetre
mM	Millimolar
Mw	Molecular weight
NaCl	Sodium chloride
Na_3PO_4	Sodium phosphate
ng	Nanogram
$(NH_4)_2SO_4$	Ammonium sulfate
nm	Nanometre

NMR	NMR
OD	Optical density
OPDA	Oxophytodienoic acid
OPR	12-oxophytodienoate reductase
ORF	Open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PMA	Phorbol 12-myristate-13-acetate
PE	Phorbol ester
PEG-6000	Polyethylene Glycol 6000
pI	Isoelectric point
PIC	Polymorphic information content
PKC	Protein kinase C
pmol	Picomole
PPB	Percentage of polymorphic bands
qRT-PCR	Quantitative real-time polymerase chain reaction
RAPD	Random amplification of polymorphic DNA
RE	Restriction enzyme
RIP	Ribosome inactivating protein
R/L	Restriction digestion/ligation
RNA	Ribonucleic acid
rpm	Revolutions per minute
s	Second(s)
SDS	Sodium dodecyl sulfate
SA	Salicylic acid
SNP	Single nucleotide polymorphism)
SSR	Simple sequence repeat
T_a	Annealing temperature
Tris	2-amino-2-hydroxymethyl-1,3-propanediol
U	Units
UPGMA	Unweighted pair group method with arithmetic mean)
USA	United States of America
USD	U
UTR	Untranslated region

v/v	Volume to volume ratio
w/v	Weight to volume ratio

References

ACHTEN, W. M. J., VERCHOT, L., FRANKEN, Y. J., MATHIJS, E., SINGH, V. P., AERTS, R. & MUYS, B. 2008. *Jatropha* bio-diesel production and use. *Biomass & Bioenergy,* 32**,** 1063-1084.

ACOSTA, I. F. & FARMER, E. E. 2010. Jasmonates. *The Arabidopsis Book,* 8**,** 1-13.

ACQUAAH, G. 2007. *Principles of plant genetics and breeding,* Wiley-Blackwell.

ADERIBIGBE, A., JOHNSON, C., MAKKAR, H., BECKER, K. & FOIDL, N. 1997. Chemical composition and effect of heat on organic matter and nitrogen degradability and some antinutritional components of *Jatropha* meal. *Animal feed science and technology,* 67**,** 223-243.

AJMONE-MARSAN, P., GORNI, C., CHITTO, A., REDAELLI, R., VAN VIJK, R., STAM, P. & MOTTO, M. 2001. Identification of QTLs for grain yield and grain-related traits of maize (*Zeamays* L.) using an AFLP map, different testers, and cofactor analysis. *Theoretical and Applied Genetics,* 102**,** 230-243.

AJMONE-MARSAN, P., VALENTINI, A., CASSANDRO, M., VECCHIOTTI-ANTALDI, G., BERTONI, G. & KUIPER, M. 1997. AFLP^{TM} markers for DNA fingerprinting in cattle. *Animal Genetics,* 28**,** 418-426.

ALONSO-BLANCO, C., PEETERS, A. J., KOORNNEEF, M., LISTER, C., DEAN, C., VAN DEN BOSCH, N., POT, J. & KUIPER, M. T. 1998. Development of an AFLP based linkage map of Ler, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a Ler/Cvi recombinant inbred line population. *Plant J,* 14**,** 259-271.

AMBROSI, D. G., GALLA, G., PURELLI, M., BARBI, T., FABBRI, A., LUCRETTI, S., SHARBEL, T. F. & BARCACCIA, G. 2010. DNA markers and FCSS analyses shed light on the genetic diversity and reproductive strategy of *Jatropha curcas* L. *Diversity,* 2**,** 810-836.

ANWAR, F., RASHID, U., JAMIL, A. & BHATTI, H. N. 2010. *Jatropha curcas* seed oil as a viable source for biodiesel. *Pakistan Journal of Botany,* 42**,** 575-582.

BAMGBOYE, A. & HANSEN, A. 2008. Prediction of cetane number of biodiesel fuel from the fatty acid methyl ester (FAME) composition. *International Agrophysics,* 22**,** 21.

BARBIERI, L., BATTELLI, M. G. & STIRPE, F. 1993. Ribosome-inactivating proteins from plants. *Biochim Biophys Acta,* 1154**,** 237-282.

BARBIERI, L., VALBONESI, P., BONORA, E., GORINI, P., BOLOGNESI, A. & STIRPE, F. 1997. Polynucleotide:adenosine glycosidase activity of ribosome-inactivating proteins: effect on DNA, RNA and poly(A). *Nucleic Acids Research,* 25**,** 518-522.

BARI, R. & JONES, J. D. 2009. Role of plant hormones in plant defence responses. *Plant Molecular Biology,* 69**,** 473-488.

BASHA, S. D., FRANCIS, G., MAKKAR, H. P. S., BECKER, K. & SUJATHA, M. 2009. A comparative study of biochemical traits and molecular markers for assessment of genetic relationships between Jatropha curcas L. germplasm from different countries. *Plant Science,* 176**,** 812-823.

BASHA, S. D. & SUJATHA, M. 2007. Inter and intra-population variability of *Jatropha curcas* (L.) characterized by RAPD and ISSR markers and development of population-specific SCAR markers. *Euphytica,* 156**,** 375-386.

BASHA, S. D. & SUJATHA, M. 2009. Genetic analysis of *Jatropha* species and interspecific hybrids of Jatropha curcas using nuclear and organelle specific markers. *Euphytica,* 168**,** 197-214.

BASZCZYK, L., TYRKA, M. & CHEKOWSKI, J. 2005. AFLP based markers for leaf rust resistance genes in common wheat. *Journal of Applied Genetics,* 46**,** 357-364.

BENSCH, S. & AKESSON, M. 2005. Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology,* 14**,** 2899-2914.

BEUTLER, E., GELBART, T., HAN, J. H., KOZIOL, J. A. & BEUTLER, B. 1989. Evolution of the genome and the genetic code: selection at the dinucleotide level by methylation and polyribonucleotide cleavage. *Proceedings of the National Academy of Sciences of the United States of America,* 86**,** 192-196.

BRADFORD, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry,* 72**,** 248-254.

BRITISH PETROLEUM 2010. BP Statistical Review of World Energy June 2010. *BP Statistical Review of World Energy*.

BRITISH PETROLEUM 2011. BP Energy Outlook 2030.

BRITTAINE, R. & LUTALADIO, N. 2010. *Jatropha*: a smallholder bioenergy crop. *Integrated Crop Management,* 8.

BRUNAK, S., ENGELBRECHT, J. & KNUDSEN, S. 1991. Prediction of human mRNA donor and acceptor sites from the DNA sequence. *Journal of Molecular Biology,* 220**,** 49-65.

BUCHANAN, B. B., GRUISSEM, W. & JONES, R. L. 2000. *Biochemistry & Molecular Biology of Plants,* American Society of Plant Physiologists Rockville, MD.

CARVALHO, C. R., CLARINDO, W. R., PRACA, M. M., ARAUJO, F. S. & CARELS, N. 2008. Genome size, base composition and karyotype of *Jatropha curcas* L., an important biofuel plant. *Plant Science,* 174**,** 613-617.

CHEONG, J. J. & CHOI, Y. D. 2003. Methyl jasmonate as a vital substance in plants. *Trends in Genetics,* 19**,** 409-413.

CHINI, A., BOTER, M. & SOLANO, R. 2009. Plant oxylipins: COI1/JAZs/MYC2 as the core jasmonic acid signalling module. *The FEBS Journal,* 276**,** 4682-4692.

COLBERT, T., TILL, B. J., TOMPA, R., REYNOLDS, S., STEINE, M. N., YEUNG, A. T., MCCALLUM, C. M., COMAI, L. & HENIKOFF, S. 2001. High-throughput screening for induced point mutations. *Plant Physiology,* 126**,** 480-484.

COOPER, D. N. & YOUSSOUFIAN, H. 1988. The CpG dinucleotide and human genetic disease. *Human Genetics,* 78**,** 151-155.

COOPER, J. L., TILL, B. J., LAPORT, R. G., DARLOW, M. C., KLEFFNER, J. M., JAMAI, A., EL-MELLOUKI, T., LIU, S., RITCHIE, R., NIELSEN, N., BILYEU, K. D., MEKSEM, K., COMAI, L. & HENIKOFF, S. 2008. TILLING to detect induced mutations in soybean. *BMC Plant Biology,* 8**,** 9.

COSTA, G. G., CARDOSO, K. C., DEL BEM, L. E., LIMA, A. C., CUNHA, M. A., DE CAMPOS-LEITE, L., VICENTINI, R., PAPES, F., MOREIRA, R. C., YUNES, J. A., CAMPOS, F. A. & DA SILVA, M. J. 2010. Transcriptome analysis of the oil-rich seed of the bioenergy crop *Jatropha curcas* L. *BMC Genomics,* 11**,** 462.

COX, T. S., GLOVER, J. D., VAN TASSEL, D. L., COX, C. M. & DEHAAN, L. R. 2006. Prospects for developing perennial-grain crops. *Bioscience,* 56**,** 649-659.

DAVE, A., HERNANDEZ, M. L., HE, Z., ANDRIOTIS, V. M., VAISTIJ, F. E., LARSON, T. R. & GRAHAM, I. A. 2011. 12-oxo-phytodienoic acid accumulation during seed development represses seed germination in *Arabidopsis*. *Plant Cell*.

DE FREITAS, L. C. & KANEKO, S. 2011. Ethanol demand under the flex-fuel technology regime in Brazil. *Energy Economics*.

DEVAPPA, R. K., MAKKAR, H. P. & BECKER, K. 2010a. *Jatropha* toxicity—a review. *Journal of Toxicol Environmental Health, Part B,* 13**,** 476-507.

DEVAPPA, R. K., MAKKAR, H. P. S. & BECKER, K. 2010b. Biodegradation of *Jatropha curcas* phorbol esters in soil. *Journal of the Science of Food and Agriculture,* 90**,** 2090-2097.

DIVAKARA, B. N., UPADHYAYA, H. D., WANI, S. P. & GOWDA, C. L. L. 2010. Biology and genetic improvement of *Jatropha curcas* L.: a review. *Applied Energy,* 87**,** 732-742.

DUNN, R., SHOCKLEY, M. & BAGBY, M. 1996. Improving the low-temperature properties of alternative diesel fuels: vegetable oil-derived methyl esters. *Journal of the American Oil Chemists' Society,* 73**,** 1719-1728.

EASTMOND, P. J. & GRAHAM, I. A. 2001. Re-examining the role of the glyoxylate cycle in oilseeds. *Trends in Plant Science,* 6**,** 72-77.

EFRON, B., HALLORAN, E. & HOLMES, S. 1996. Bootstrap confidence levels for phylogenetic trees. *Proceedings of the National Academy of Sciences of the United States of America,* 93**,** 7085-7090.

ELGAVISH, S. & SHAANAN, B. 1997. Lectin-carbohydrate interactions: different folds, common recognition principles. *Trends in Biochemical Sciences,* 22**,** 462-467.

EVANNO, G., REGNAUT, S. & GOUDET, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology,* 14**,** 2611-2620.

FARMER, E. E. & RYAN, C. A. 1990. Interplant communication: airborne methyl jasmonate induces synthesis of proteinase inhibitors in plant leaves. *Proceedings of the National Academy of Sciences of the United States of America,* 87**,** 7713.

FARMER, E. E. & RYAN, C. A. 1992. Octadecanoid precursors of jasmonic acid activate the synthesis of wound-inducible proteinase inhibitors. *The Plant Cell,* 4**,** 129.

FELSENSTEIN, J. & KISHINO, H. 1993. Is there something wrong with the Bootstrap on phylogenies? A reply to Hillis and Bull. *Systematic Biology,* 42**,** 193-200.

FINNEGAN, E. J., GENGER, R. K., PEACOCK, W. J. & DENNIS, E. S. 1998. DNA methylation in plants. *Annual Review of Plant Physiology and Plant Molecular Biology,* 49**,** 223-247.

FOIDL, N., FOIDL, G., SANCHEZ, M., MITTELBACH, M. & HACKEL, S. 1996. *Jatropha curcas* L. as a source for the production of biofuel in Nicaragua. *Bioresource Technology,* 58**,** 77-82.

GASIC, K., HERNANDEZ, A. & KORBAN, S. S. 2004. RNA extraction from different apple tissues rich in polyphenols and polysaccharides for cDNA library construction. *Urbana,* 51**,** 61801.

GEPTS, P. 2006. Plant genetic resources conservation and utilization: The accomplishments and future of a societal insurance policy. *Crop Science,* 46**,** 2278-2292.

GHOSH, A., CHIKARA, J., CHAUDHARY, D. R., PRAKASH, A. R., BORICHA, G. & ZALA, A. 2010. Paclobutrazol arrests vegetative growth and unveils unexpressed yield potential of *Jatropha curcas*. *Journal of Plant Growth Regulation,* 29**,** 307-315.

GHOSH, A., PATOLIA, J., CHAUDHARRY, D., CHIKARA, J., RAO, S., KUMAR, D., BORICHA, G. & ZALA, A. 2007. Response of *Jatropha curcas* under different spacing to *Jatropha* de-oiled cake.

GLAUSER, G., GRATA, E., DUBUGNON, L., RUDAZ, S., FARMER, E. E. & WOLFENDER, J. L. 2008. Spatial and temporal dynamics of jasmonate synthesis and accumulation in *Arabidopsis* in response to wounding. *Journal of Biological Chemistry,* 283**,** 16400.

GLOVER, J. D., COX, C. M. & REGANOLD, J. P. 2007. Future farming: a return to roots? *Scientific American,* 297**,** 82-89.

GOEL, G., MAKKAR, H. P., FRANCIS, G. & BECKER, K. 2007. Phorbol esters: structure, biological activity, and toxicity in animals. *International Journal of Toxicology,* 26**,** 279-288.

GRAHAM, I. A. 2008. Seed storage oil mobilization. *Annual Review of Plant Biology,* 59**,** 115-142.

GUBITZ, G., MITTELBACH, M. & TRABI, M. 1999. Exploitation of the tropical oil seed plant *Jatropha curcas* L. *Bioresource Technology,* 67**,** 73-82.

HAAS, W., STERK, H. & MITTELBACH, M. 2002. Novel 12-deoxy-16-hydroxyphorbol diesters isolated from the seed oil of *Jatropha curcas*. *Journal of Natural Products,* 65**,** 1434-1440.

HALL, B. G. 2004. *Phylogenetic trees made easy: a how-to manual,* Sinauer Associates.

HARRIS, H. C., MCWILLIAM, J. R. & MASON, W. K. 1978. Influence of temperature on oil content and composition of sunflower seed. *Australian Journal of Agricultural Research,* 29**,** 1203-1212.

HARTLEY, M. R. & LORD, J. M. 2004a. Cytotoxic ribosome-inactivating lectins from plants. *Biochim Biophys Acta,* 1701**,** 1-14.

HARTLEY, M. R. & LORD, J. M. 2004b. Genetics of ribosome-inactivating proteins. *Mini Reviews in Medicinal Chemistry,* 4**,** 487-492.

HE, W., KING, A. J., KHAN, A. M., CUEVAS, J. A., RAMIARAMANANA, D. & GRAHAM, I. A. 2011. Analysis of seed phorbol-ester and curcin content together with genetic diversity in multiple provenances of *Jatropha curcas* (L.) from Madagascar and Mexico. *Plant Physiology and Biochemistry,* 49**,** 1183-1190.

HEBSGAARD, S. M., KORNING, P. G., TOLSTRUP, N., ENGELBRECHT, J., ROUZ , P. & BRUNAK, S. 1996. Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information. *Nucleic Acids Research,* 24**,** 3439.

HELLER, J. 1996. *Physic Nut. Jatropha curcas L. Promoting the conservation and use of underutilized and negleted crops.,* Institure of Plant Genetics and Crop Research, Gatersleben, Germany & International Plant Genetic Resource Institute, Rome.

HELMY, M., LOMBARD, S. & PIERONI, G. 1999. Ricin RCA60: evidence of its phospholipase activity. *Biochemical and Biophysical Research Communications,* 258**,** 252-255.

HOLDER, M. & LEWIS, P. O. 2003. Phylogeny estimation: traditional and Bayesian approaches. *Nature Reviews Genetics,* 4**,** 275-284.

HOWE, G. A. & JANDER, G. 2008. Plant immunity to insect herbivores. *Annual Review of Plant Biology,* 59**,** 41-66.

HOWE, G. A., LEE, G. I., ITOH, A., LI, L. & DEROCHER, A. E. 2000. Cytochrome P450-dependent metabolism of oxylipins in tomato. Cloning and expression of allene oxide synthase and fatty acid hydroperoxide lyase. *Plant Physiology,* 123**,** 711-724.

HUANG, M. X., HOU, P., WEI, Q., XU, Y. & CHEN, F. 2008. A ribosome-inactivating protein (curcin 2) induced from *Jatropha curcas* can reduce viral and fungal infection in transgenic tobacco. *Plant Growth Regulation,* 54**,** 115-123.

ICNGUR, Y. & ALTIPARMAK, D. 2003. Effect of fuel cetane number and injection pressure on a DI Diesel engine performance and emissions. *Energy conversion and management,* 44**,** 389-397.

IGLESIAS, R., PEREZ, Y., DE TORRE, C., FERRERAS, J. M., ANTOLIN, P., JIMENEZ, P., ROJO, M. A., MENDEZ, E. & GIRBES, T. 2005. Molecular characterization and systemic induction of single-chain ribosome-inactivating proteins (RIPs) in sugar beet (*Beta vulgaris*) leaves. *Journal of Experimental Botany,* 56**,** 1675-1684.

IMAHARA, H., MINAMI, E. & SAKA, S. 2006. Thermodynamic study on cloud point of biodiesel with its fatty acid composition. *Fuel,* 85**,** 1666-1670.

INTERNATIONAL ATOMIC ENERGY AGENCY 2009. IEAE annual report.

IPEK, M., IPEK, A., ALMQUIST, S. G. & SIMON, P. W. 2005. Demonstration of linkage and development of the first low-density genetic map of garlic, based on AFLP markers. *Theorotical and Applied Genetics,* 110**,** 228-236.

JONES, C. J., EDWARDS, K. J., CASTAGLIONE, S., WINFIELD, M. O., SALA, F., VAN DE WIEL, C., BREDEMEIJER, G., VOSMAN, B., MATTHES, M., DALY, A., BRETTSCHNEIDER, R., BETTINI, P., BUIATTI, M., MAESTRI, E., MALCEVSCHI, A., MARMIROLI, N., AERT, R., VOLCKAERT, G., RUEDA, J., LINACERO, R., VAZQUEZ, A. & KARP, A. 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding,* 3**,** 381-390.

JONGSCHAAP, R., CORR , W., BINDRABAN, P. & BRANDENBURG, W. 2007. Claims and Facts on *Jatropha curcas* L. *Wageningen: Plant Research International*.

JOSHI, C., MATHUR, P. & KHARE, S. K. 2011. Degradation of phorbol esters by Pseudomonas aeruginosa PseA during solid-state fermentation of deoiled *Jatropha curcas* seed cake. *Bioresource Technology,* 102**,** 4815-4819.

KAUL, S *et al*. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature,* 408**,** 796-815.

KING, A. J., HE, W., CUEVAS, J. A., FREUDENBERGER, M., RAMIARAMANANA, D. & GRAHAM, I. A. 2009. Potential of *Jatropha curcas* as a source of renewable oil and animal feed. *Journal of Experimental Botany,* 60**,** 2897-2905.

KING, A. J., LI, Y. & GRAHAM, I. A. 2010. Profiling the developing *Jatropha curcas* L. seed transcriptome by pyrosequencing. *BioEnergy Research***,** 1-11.

KNOTHE, G. 2008. "Designer" biodiesel: Optimizing fatty ester composition to improve fuel properties. *Energy & Fuels,* 22**,** 1358-1364.

KRAUSS, S. L. 2000. Accurate gene diversity estimates from amplified fragment length polymorphism (AFLP) markers. *Molecular Ecology,* 9**,** 1241-1245.

KRISHNAKUMAR, J., VENKATACHALAPATHY, V. S. K. & ELANCHELIYAN, S. 2008. Technical aspects of biodiesel production from vegetable oils. *Thermal Science,* 12**,** 159-169.

KUMAR, V., MAKKAR, H. P. S., DEVAPPA, R. K. & BECKER, K. 2011. Isolation of phytate from *Jatropha curcas* kernel meal and effects of isolated phytate on growth, digestive physiology and metabolic changes in Nile tilapia (*Oreochromis niloticus* L.). *Food and Chemical Toxicology,* 49**,** 2144-2156.

LAEMMLI, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* 227**,** 680-685.

LANTERI, S., SABA, E., CADINU, M., MALLICA, G. M., BAGHINO, L. & PORTIS, E. 2004. Amplified fragment length polymorphism for genetic diversity assessment in globe artichoke. *Theorotical and Applied Genetics,* 108**,** 1534-1544.

LARSON, T. R. & GRAHAM, I. A. 2001. Technical Advance: a novel technique for the sensitive quantification of acyl CoA esters from plant tissues. *The Plant Journal,* 25**,** 115-125.

LEE, I., JOHNSON, L. A. & HAMMOND, E. G. 1995. Use of branched-chain esters to reduce the crystallization temperature of biodiesel. *Journal of the American Oil Chemists' Society,* 72**,** 1155-1160.

LEE, I., JOHNSON, L. A. & HAMMOND, E. G. 1996. Reducing the crystallization temperature of biodiesel by winterizing methyl soyate. *Journal of the American Oil Chemists' Society,* 73**,** 631-636.

LI, M. X., YEUNG, H. W., PAN, L. P. & CHAN, S. I. 1991. Trichosanthin, a potent HIV-1 inhibitor, can cleave supercoiled DNA in vitro. *Nucleic Acids Research,* 19**,** 6309.

LIKENS, G. E. & BORMANN, F. H. 1974. Acid rain: a serious regional environmental problem. *Science,* 184**,** 1176.

LIN, J., CHEN, Y., XU, Y., YAN, F., TANG, L. & CHEN, F. 2003a. Cloning and expression of curcin, a ribosome inactivating protein from the seeds of *Jatropha curcas*. *Acta Botanica Sinica,* 45**,** 858-863.

LIN, J., LI, Y. X., ZHOU, X. W., TANG, K. X. & CHEN, F. 2003b. Cloning and characterization of a curcin gene encoding a ribosome inactivating protein from *Jatropha curcas*. *DNA Sequence,* 14**,** 311-317.

LIN, J., ZHOU, X., WANG, J., JIANG, P. & TANG, K. 2010. Purification and characterization of curcin, a toxic lectin from the seed of Jatropha curcas. *Prep Biochem Biotechnol,* 40**,** 107-18.

LIONNETON, E., RAVERA, S., SANCHEZ, L., AUBERT, G., DELOURME, R. & OCHATT, S. 2002. Development of an AFLP-based linkage map and localization of QTLs for seed fatty acid content in condiment mustard (*Brassica juncea*). *Genome,* 45**,** 1203-1215.

LIU, K. & MUSE, S. V. 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics,* 21**,** 2128-2129.

MAKKAR, H., FRANCIS, G. & BECKER, K. 2007. Bioactivity of phytochemicals in some lesser-known plants and their effects and potential applications in livestock and aquaculture production systems. *Animal,* 1**,** 1371-1391.

MAKKAR, H. P. & BECKER, K. 1999. Nutritional studies on rats and fish (carp *Cyprinus carpio*) fed diets containing unheated and heated *Jatropha curcas* meal of a non-toxic provenance. *Plant Foods for Human Nutrition,* 53**,** 183-192.

MAKKAR, H. P., BECKER, K. & SCHMOOK, B. 1998a. Edible provenances of *Jatropha curcas* from Quintana Roo state of Mexico and effect of roasting on antinutrient and toxic factors in seeds. *Plant Foods for Human Nutrition,* 52**,** 31-36.

MAKKAR, H. P. S., ADERIBIGBE, A. O. & BECKER, K. 1998b. Comparative evaluation of non-toxic and toxic varieties of *Jatropha curcas* for chemical composition, digestibility, protein degradability and toxic factors. *Food Chemistry,* 62**,** 207-215.

MAKKAR, H. P. S. & BECKER, K. 2009. *Jatropha curcas,* a promising crop for the generation of biodiesel and value added coproducts. *European Journal of Lipid Science and Technology,* 111**,** 773-787.

MARRIOTT, K. M. & NORTHCOTE, D. H. 1975. The breakdown of lipid reserves in the endosperm of germinating castor beans. *Biochem J,* 148**,** 139-44.

MARTINEZ-HERRERA, J., SIDDHURAJU, P., FRANCIS, G., DAVILA-ORTIZ, G. & BECKER, K. 2006. Chemical composition, toxic/antimetabolic constituents, and effects of different treatments on their levels, in four provenances of *Jatropha curcas* L. from Mexico. *Food Chemistry,* 96**,** 80-89.

MAU, B., NEWTON, M. A. & LARGET, B. 1999. Bayesian phylogenetic inference via Markov chain Monte Carlo methods. *Biometrics,* 55**,** 1-12.

MCLENACHAN, P. A., STOCKLER, K., WINKWORTH, R. C., MCBREEN, K., ZAUNER, S. & LOCKHART, P. J. 2000. Markers derived from amplified fragment length polymorphism gels for plant ecology and evolution studies. *Molecular Ecology,* 9**,** 1899-1903.

MEUDT, H. M. & CLARKE, A. C. 2007. Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends in Plant Science,* 12**,** 106-117.

MOCK, J., NG, T., WONG, R., YAO, Q., YEUNG, H. & FONG, W. 1996. Demonstration of ribonuclease activity in the plant ribosome-inactivating proteins alpha- and beta-momorcharins. *Life Sciences,* 59**,** 1853-1859.

MOROZOVA, O. & MARRA, M. 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics,* 92**,** 255-264.

MUTO, S. & BEEVERS, H. 1974. Lipase activities in castor bean endosperm during germination. *Plant Physiology,* 54**,** 23.

MYBURG, A., REMINGTON, D., O¡¯MALLEY, D., SEDEROFF, R. & WHETTEN, R. 2001. Research report high-throughput AFLP analysis using infrared dye-labeled primers and an automated DNA sequencer. *Biotechniques,* 30**,** 348-357.

NATARAJAN, P., KANAGASABAPATHY, D., GUNADAYALAN, G., PANCHALINGAM, J., SHREE, N., SUGANTHAM, P. A., SINGH, K. K. & MADASAMY, P. 2010. Gene discovery from *Jatropha curcas* by sequencing of ESTs from normalized and full-length enriched cDNA library from developing seeds. *BMC Genomics,* 11**,** 606.

NATARAJAN, P. & PARANI, M. 2011. De novo assembly and transcriptome analysis of five major tissues of *Jatropha curcas* L. using GS FLX titanium platform of 454 pyrosequencing. *BMC Genomics,* 12**,** 191.

NEI, M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America,* 70**,** 3321-3323.

NICOLAS, E., BEGGS, J. M., HALTIWANGER, B. M. & TARASCHI, T. F. 1997. Direct evidence for the deoxyribonuclease activity of the plant ribosome inactivating protein gelonin. *FEBS letters,* 406**,** 162-164.

NIELSEN, K. & BOSTON, R. S. 2001. Ribosome-inactivating proteins: a plant perspective. *Annual Review of Plant Physiology and Plant Molecular Biology,* 52**,** 785-816.

OSSOWSKI, S., SCHNEEBERGER, K., LUCAS-LLEDO, J. I., WARTHMANN, N., CLARK, R. M., SHAW, R. G., WEIGEL, D. & LYNCH, M. 2010. The Rate and Molecular Spectrum of Spontaneous Mutations in *Arabidopsis thaliana*. *Science,* 327**,** 92-94.

PAGLIARO, M., CIRIMINNA, R., KIMURA, H., ROSSI, M. & DELLA PINA, C. 2007. From glycerol to value-added products. *Angewandte Chemie-International Edition,* 46**,** 4434-4440.

PANDEY, A. 2009. *Handbook of plant-based biofuels,* CRC.

PARK, S. W., VEPACHEDU, R., SHARMA, N. & VIVANCO, J. M. 2004. Ribosome-inactivating proteins in plant biology. *Planta,* 219**,** 1093-1096.

PENFIELD, S., LI, Y., GILDAY, A. D., GRAHAM, S. & GRAHAM, I. A. 2006. *Arabidopsis* ABA INSENSITIVE4 regulates lipid mobilization in the embryo and reveals repression of seed germination by the endosperm. *The Plant Cell,* 18**,** 1887.

PENFIELD, S., RYLOTT, E. L., GILDAY, A. D., GRAHAM, S., LARSON, T. R. & GRAHAM, I. A. 2004. Reserve mobilization in the *Arabidopsis* endosperm fuels hypocotyl elongation in the dark, is independent of abscisic acid, and requires PHOSPHOENOLPYRUVATE CARBOXYKINASE1. *The Plant Cell,* 16**,** 2705.

PEUMANS, W. J., HAO, Q. & VAN DAMME, E. J. 2001. Ribosome-inactivating proteins from plants: more than RNA N-glycosidases? *The FASEB Journal,* 15**,** 1493-1506.

POPLUECHAI, S., BREVIARIO, D., MULPURI, S., MAKKAR, H. P. S., RAORANE, M., REDDY, A. R., PALCHETTI, E., GATEHOUSE, A. M. R., SYERS, K. J. & O'DONNELL, A. G. 2009. Narrow genetic and apparent phenetic diversity in *Jatropha curcas*: initial success with generating low phorbol ester interspecific hybrids. Nature Precedings.

POWELL, W., MORGANTE, M., ANDRE, C., HANAFEY, M., VOGEL, J., TINGEY, S. & RAFALSKI, A. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding,* 2**,** 225-238.

POYET, J. L. & HOEVELER, A. 1997. Presence of an intron in a gene of PAP II, the ribosome-inactivating protein from summer leaves of *Phytolacca americana*. *Annals of Botany,* 80**,** 685-688.

PRITCHARD, J. K., STEPHENS, M. & DONNELLY, P. 2000. Inference of population structure using multilocus genotype data. *Genetics,* 155**,** 945-959.

QIN, X., SHAO, C., HOU, P., GAO, J., LEI, N., JIANG, L., YE, S., GOU, C., LUO, S., ZHENG, X., GU, X., ZHU, X., XU, Y. & CHEN, F. 2010. Different functions and expression profiles of curcin and curcin-L in *Jatropha curcas* L. *Z Naturforsch C,* 65**,** 355-362.

QIN, X., ZHANG, J., SHAO, C., LIN, S., JIANG, L., ZHANG, S., XU, Y. & CHEN, F. 2009a. Isolation and characterization of a curcin promoter from *Jatropha curcas* L. and Its regulation of gene expression in transgenic tobacco plants. *Plant Molecular Biology Reporter,* 27**,** 275-281.

QIN, X., ZHENG, X., SHAO, C., GAO, J., JIANG, L., ZHU, X., YAN, F., TANG, L., XU, Y. & CHEN, F. 2009b. Stress-induced curcin-L promoter in leaves of *Jatropha curcas* L. and characterization in transgenic tobacco. *Planta,* 230**,** 387-395.

RANNALA, B. & YANG, Z. 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *Journal of Molecular Evolution,* 43**,** 304-311.

REINBOTHE, S., MOLLENHAUER, B. & REINBOTHE, C. 1994a. JIPs and RIPs: the regulation of plant gene expression by jasmonates in response to environmental cues and pathogens. *Plant Cell,* 6**,** 1197-1209.

REINBOTHE, S., REINBOTHE, C., LEHMANN, J., BECKER, W., APEL, K. & PARTHIER, B. 1994b. JIP60, a methyl jasmonate-induced ribosome-inactivating protein involved in plant stress reactions. *Proceedings of the National Academy of Sciences of the United States of America,* 91**,** 7012-7126.

RENNER, A., ZELT, T. & GERTEISER, S. 2008. Global Market Study on *Jatropha*. *Final Report Prepared for the World Wildlife Fund for Nature (WWF)(GEXSI, London).*

REYMOND, P., WEBER, H., DAMOND, M. & FARMER, E. E. 2000. Differential gene expression in response to mechanical wounding and insect feeding in *Arabidopsis*. *Plant Cell,* 12**,** 707-720.

RIPPMANN, J. F., MICHALOWSKI, C. B., NELSON, D. E. & BOHNERT, H. J. 1997. Induction of a ribosome-inactivating protein upon environmental stress. *Plant Molecular Biology,* 35**,** 701-709.

ROLDAN-RUIZ, I., DENDAUW, J., VAN BOCKSTAELE, E., DEPICKER, A. & DE LOOSE, M. 2000. AFLP markers reveal high polymorphic rates in ryegrasses (*Lolium spp.*). *Molecular Breeding,* 6**,** 125-134.

RONCUZZI, L. & GASPERI-CAMPANI, A. 1996. DNA-nuclease activity of the single-chain ribosome-inactivating proteins dianthin 30, saporin 6 and gelonin. *FEBS letters,* 392**,** 16-20.

ROSADO, T. B., LAVIOLA, B. G., FARIA, D. A., PAPPAS, M. R., BHERING, L. L., QUIRINO, B. & GRATTAPAGLIA, D. 2010. Molecular markers reveal limited genetic diversity in a large germplasm collection of the biofuel crop *Jatropha curcas* L. in Brazil. *Crop Science,* 50**,** 2372-2382.

ROSENBERG, N. 2004. Distruct: a program for the graphical display of structure results. *Molecular Ecology Notes,* 4**,** 137-138.

ROTHBERG, J. & LEAMON, J. 2008. The development and impact of 454 sequencing. *Nature biotechnology,* 26**,** 1117-1124.

RYLOTT, E., HOOKS, M. & GRAHAM, I. 2001. Co-ordinate regulation of genes involved in storage lipid mobilization in *Arabidopsis thaliana*. *Biochemical Society Transactions,* 29**,** 283-286.

SARIN, R., SHARMA, M., SINHARAY, S. & MALHOTRA, R. K. 2007. *Jatropha*-palm biodiesel blends: an optimum mix for Asia. *Fuel,* 86**,** 1365-1371.

SATO, S., HIRAKAWA, H., ISOBE, S., FUKAI, E., WATANABE, A., KATO, M., KAWASHIMA, K., MINAMI, C., MURAKI, A., NAKAZAKI, N., TAKAHASHI, C., NAKAYAMA, S., KISHIDA, Y., KOHARA, M.,

YAMADA, M., TSURUOKA, H., SASAMOTO, S., TABATA, S., AIZU, T., TOYODA, A., SHIN, I. T., MINAKUCHI, Y., KOHARA, Y., FUJIYAMA, A., TSUCHIMOTO, S., KAJIYAMA, S., MAKIGANO, E., OHMIDO, N., SHIBAGAKI, N., CARTAGENA, J. A., WADA, N., KOHINATA, T., ATEFEH, A., YUASA, S., MATSUNAGA, S. & FUKUI, K. 2011. Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. *DNA Research,* 18**,** 65-76.

SCHIERHOLT, A., BECKER, H. & ECKE, W. 2000. Mapping a high oleic acid mutation in winter oilseed rape (*Brassica napus* L.). *Theoretical and Applied Genetics,* 101**,** 897-901.

SCHMOOK, B. & SERRALTA-PERAZA, L. Year. *Jatropha curcas*: distribution and uses in the Yucatan Peninsula. *In:* Proceedings of First International Symposium on Biofuel and Industrial Products from *Jatropha curcas* and other Tropical Oil Seed Plants, 1997 Managua, Nicaragua.

SCHNEIDER, K. 2005. *Mapping Populations and Principles of Genetic Mapping,* Vch Verlagsgesellschaft Mbh.

SEINFELD, J. H., PANDIS, S. N. & KNOVEL 1998. *Atmospheric chemistry and physics: from air pollution to climate change,* Wiley New York.

SHARMA, N., PARK, S. W., VEPACHEDU, R., BARBIERI, L., CIANI, M., STIRPE, F., SAVARY, B. J. & VIVANCO, J. M. 2004. Isolation and characterization of an RIP (ribosome-inactivating protein) like protein from tobacco with dual enzymatic activity. *Plant Physiology,* 134**,** 171.

SHEN, J. L., JIA, X. N., NI, H. Q., SUN, P. G., NIU, S. H. & CHEN, X. Y. 2010. AFLP analysis of genetic diversity of *Jatropha curcas* grown in Hainan, China. *Trees-Structure and Function,* 24**,** 455-462.

SHIH, N. R., MCDONALD, K. A., JACKMAN, A. P., GIRB S, T. & IGLESIAS, R. 1997. Bifunctional plant defence enzymes with chitinase and ribosome inactivating activities from Trichosanthes kirilowii cell cultures. *Plant Science,* 130**,** 145-150.

SLADE, A. J., FUERSTENBERG, S. I., LOEFFLER, D., STEINE, M. N. & FACCIOTTI, D. 2005. A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. *Nature biotechnology,* 23**,** 75-81.

SOAVE, C. & SALAMINI, F. 1984. Organization and regulation of zein genes in maize endosperm. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences,* 304**,** 341-347.

SOH, A., WONG, G., HOR, T., TAN, C. & CHEW, P. 2003. Oil palm genetic improvement. *Plant Breeding Reviews,* 22.

SOLOMON, S., QIN, D., MANNING, M., CHEN, Z., MARQUIS, M., AVERYT, K. B., TIGNOR, M. & MILLER, H. L. 2007. *Climate change 2007: the physical science basis: contribution of working group I to the fourth assessment report of the Intergovernmental Panel on Climate Change,* Cambridge University Press.

SONG, S. K., CHOI, Y., MOON, Y. H., KIM, S. G., CHOI, Y. D. & LEE, J. S. 2000. Systemic induction of a *Phytolacca insularis* antiviral protein gene by mechanical wounding, jasmonic acid, and abscisic acid. *Plant Molecular Biology,* 43**,** 439-450.

SRIVASTAVA, A. & PRASAD, R. 2000. Triglycerides-based diesel fuels. *Renewable & Sustainable Energy Reviews,* 4**,** 111-133.

STENZEL, I., HAUSE, B., FEUSSNER, I. & WASTERNACK, C. 2003a. Transcriptional activation of jasmonate biosynthesis enzymes is not reflected at protein level. *Advanced Research on Plant Lipids***,** 267-270.

STENZEL, I., HAUSE, B., MIERSCH, O., KURZ, T., MAUCHER, H., WEICHERT, H., ZIEGLER, J., FEUSSNER, I. & WASTERNACK, C. 2003b. Jasmonate biosynthesis and the allene oxide cyclase family of *Arabidopsis thaliana*. *Plant Molecular Biology,* 51**,** 895-911.

STIRPE, F. 2004. Ribosome-inactivating proteins. *Toxicon,* 44**,** 371-383.

STIRPE, F., BAILEY, S., MILLER, S. P. & BODLEY, J. W. 1988. Modification of ribosomal RNA by ribosome-inactivating proteins from plants. *Nucleic Acids Research,* 16**,** 1349-1357.

STIRPE, F., BARBIERI, L., GORINI, P., VALBONESI, P., BOLOGNESI, A. & POLITO, L. 1996. Activities associated with the presence of ribosome-inactivating proteins increase in senescent and stressed leaves. *FEBS Letters,* 382**,** 309-312.

STIRPE, F. & BATTELLI, M. G. 2006. Ribosome-inactivating proteins: progress and problems. *Cellular and Molecular Life Sciences,* 63**,** 1850-1866.

STIRPE, F., PESSIONBRIZZI, A., LORENZONI, E., STROCCHI, P., MONTANARO, L. & SPERTI, S. 1976. Studies on proteins from seeds of *Croton tiglium* and of *Jatropha curcas* - toxic properties and inhibition of protein-synthesis *in vitro*. *Biochemical Journal,* 156**,** 1-6.

SUDHEER, P. D. V. N., MASTAN, S. G., RAHMAN, H. & REDDY, M. P. 2010a. Molecular characterization and genetic diversity analysis of *Jatropha curcas* L. in India using RAPD and AFLP analysis. *Molecular Biology Reports,* 37**,** 2249-2257.

SUDHEER, P. D. V. N., RAHMAN, H., MASTAN, S. G. & REDDY, M. P. 2010b. Isolation of novel microsatellites using FIASCO by dual probe enrichment from *Jatropha curcas* L. and study on genetic equilibrium and diversity of Indian population revealed by isolated microsatellites. *Molecular Biology Reports,* 37**,** 3785-3793.

SUJATHA, M., MAKKAR, H. P. S. & BECKER, K. 2005. Shoot bud proliferation from axillary nodes and leaf sections of non-toxic *Jatropha curcas* L. *Plant Growth Regulation,* 47**,** 83-90.

SUN, Q. B., LI, L. F., LI, Y., WU, G. J. & GE, X. J. 2008. SSR and AFLP markers reveal low genetic diversity in the biofuel plant *Jatropha curcas* in China. *Crop Science,* 48**,** 1865-1871.

SUZA, W. P., AVILA, C. A., CARRUTHERS, K., KULKARNI, S., GOGGIN, F. L. & LORENCE, A. 2010. Exploring the impact of wounding and jasmonates on ascorbate metabolism. *Plant Physiology and Biochemistry,* 48**,** 337-350.

TARTARINI, A., PITTALUGA, E., MARCOZZI, G., TESTONE, G., RODRIGUES POUSADA, R. A., GIANNINO, D. & SPAN , L. 2010. Differential expression of saporin genes upon wounding, ABA treatment and leaf development. *Physiologia Plantarum,* 140**,** 141-152.

THOMPSON, M. L., MARRIOTT, R., DOWLE, A. & GROGAN, G. 2010. Biotransformation of beta-myrcene to geraniol by a strain of *Rhodococcus erythropolis* isolated by selective enrichment from hop plants. *Applied Microbiology and Biotechnology,* 85**,** 721-730.

TILL, B. J., COOPER, J., TAI, T. H., COLOWIT, P., GREENE, E. A., HENIKOFF, S. & COMAI, L. 2007. Discovery of chemically induced mutations in rice by TILLING. *BMC Plant Biology,* 7**,** 19.

TILL, B. J., REYNOLDS, S. H., WEIL, C., SPRINGER, N., BURTNER, C., YOUNG, K., BOWERS, E., CODOMO, C. A., ENNS, L. C., ODDEN, A. R., GREENE, E. A., COMAI, L. & HENIKOFF, S. 2004. Discovery of induced point mutations in maize genes by TILLING. *BMC Plant Biology,* 4**,** 12.

TULLY, R. E. & BEEVERS, H. 1976. Protein bodies of castor bean endosperm: isolation, fractionation, and the characterization of protein components. *Plant Physiol,* 58**,** 710-716.

TURLEY, R. B. & TRELEASE, R. N. 1990. Development and regulation of 3 glyoxysomal enzymes during cotton seed maturation and growth. *Plant Molecular Biology,* 14**,** 137-146.

TYSON, K. 2009. *Biodiesel Handling and Use Guidelines,* DIANE Publishing.

VASCONCELOS, I. M. & OLIVEIRA, J. T. A. 2004. Antinutritional properties of plant lectins. *Toxicon,* 44**,** 385-403.

VOS, P., HOGERS, R., BLEEKER, M., REIJANS, M., VAN DE LEE, T., HORNES, M., FRIJTERS, A., POT, J., PELEMAN, J. & KUIPER, M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research,* 23**,** 4407-4414.

VYAS, D. & SINGH, R. 2007. Feasibility study of *Jatropha* seed husk as an open core gasifier feedstock. *Renewable Energy,* 32**,** 512-517.

WALSH, T. A., MORGAN, A. E. & HEY, T. D. 1991. Characterization and molecular cloning of a proenzyme form of a ribosome-inactivating protein from maize. *Journal of Biological Chemistry,* 266**,** 23422-23427.

WASTERNACK, C. 2007. Jasmonates: an update on biosynthesis, signal transduction and action in plant stress response, growth and development. *Annals of Botany,* 100**,** 681-697.

WASTERNACK, C. & PARTHIER, B. 1997. Jasmonate-signalled plant gene expression. *Trends in Plant Science,* 2**,** 302-307.

WASTERNACK, C., STENZEL, I., HAUSE, B., HAUSE, G., KUTTER, C., MAUCHER, H., NEUMERKEL, J., FEUSSNER, I. & MIERSCH, O. 2006. The wound response in tomato—role of jasmonic acid. *Journal of Plant Physiology,* 163**,** 297-306.

WEI, Q., HUANG, M.-X., XU, Y., ZHANG, X.-S. & CHEN, F. 2005. Expression of a ribosome inactivating protein (curcin 2) in *Jatropha curcas* is induced by stress. *Journal of Biosciences,* 30**,** 351-357.

WEIR, E. M., RIEZMAN, H., GRIENENBERGER, J. M., BECKER, W. M. & LEAVER, C. J. 1980. Regulation of glyoxysomal enzymes during germination of cucumber. *European Journal of Biochemistry,* 112**,** 469-477.

WEISING, K., NYBOM, H., WOLFF, K. & KAHL, G. 2005. *DNA fingerprinting in plants: principles, methods, and applications,* CRC press.

WOLF, R. B., CAVINS, J. F., KLEIMAN, R. & BLACK, L. T. 1982. Effect of temperature on soybean seed constituents - oil, protein, moisture, fatty-acids, amino-acids and sugars. *Journal of the American Oil Chemists Society,* 59**,** 230-232.

XIAO, J., ZHANG, H., NIU, L., WANG, X. & LU, X. 2011. Evaluation of detoxification methods on toxic and antinutritional composition and nutritional quality of proteins in *Jatropha curcas* meal. *Journal of Agricultural and Food Chemistry,* 59**,** 4040-4044.

XU, J., WANG, H. & FAN, J. 2007. Expression of a ribosome-inactivating protein gene in bitter melon is induced by *Sphaerotheca fuliginea* and abiotic stimuli. *Biotechnology Letters,* 29**,** 1605-1610.

YADAV, H., RANJAN, A., ASIF, M., MANTRI, S., SAWANT, S. & TULI, R. 2010. EST-derived SSR markers in *Jatropha curcas* L.: development, characterization, polymorphism, and transferability across the species/genera. *Tree Genetics & Genomes,* 11**,** 1-13.

YANG, C. Y., DENG, X., FANG, Z. & PENG, D. P. 2010. Selection of high-oil-yield seed sources of *Jatropha curcas* L. for biodiesel production. *Biofuels,* 1**,** 705-717.

YANG, M. F., LIU, Y. J., LIU, Y., CHEN, H., CHEN, F. & SHEN, S. H. 2009. Proteomic analysis of oil mobilization in seed germination and postgermination development of *Jatropha curcas*. *Journal of Proteome Research,* 8**,** 1441-1451.

YEH, F. & BOYLE, T. 1997. Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belgian Journal of Botany,* 129**,** 157.

YI, C. X., HONG, Y., ZHANG, S. L., LIU, X. K. & BUI, H. T. N. 2010. Does epigenetic polymorphism contribute to phenotypic variances in *Jatropha curcas* L.? *BMC Plant Biology,* 10**,** 259.

YIN, X., STAM, P., DOURLEIJN, C. J. & KROPFF, M. 1999. AFLP mapping of quantitative trait loci for yield-determining physiological characters in spring barley. *Theoretical and Applied Genetics,* 99**,** 244-253.

YIN, Z. C., WU, L. F., MAO, H. Z. & QIU, C. X. 2010. *Jatropha curcas curcin genes, tissue-specific promoters and generation of curcin-deficient transgenic Jatropha plants*.

ZARATE, S. I., KEMPEMA, L. A. & WALLING, L. L. 2007. Silverleaf whitefly induces salicylic acid defenses and suppresses effectual jasmonic acid defenses. *Plant Physiology,* 143**,** 866-875.

ZHANG, G., KAZANIETZ, M. G., BLUMBERG, P. M. & HURLEY, J. H. 1995. Crystal structure of the Cys2 activator-binding domain of protein kinase C delta in complex with phorbol ester. *Cell,* 81**,** 917-924.

ZHANG, J., QIN, X., XU, Y. & CHEN, F. 2005. Isolation and analysis on the genomic DNA sequence of members of a curcin gene-family encoding a ribosome-inactivating protein from *Jatropha curcas*. *Journal of Sichuan University (Natural Science Edition),* 42.