Global Journal of Biochemistry and Biotechnology

ISSN: xxxx-xxxx Vol. 1 (1), pp. 066-076, November, 2013. © Global Science Research Journals

Full Length Research Paper

Identification of receptor like kinase genes in coconut and development of a marker for validation of breeding materials resistant to a phytoplasma disease in Ghana

Swarbrick, P. J.[1,3], Yankey, E. N.[1,2], Nipah, J. O.[2], Quaicoe, R.[2] and Dickinson, M. J.[1]*

[1]School of Biosciences, University of Nottingham, Sutton Bonington, Leicestershire, LE12 5RD, UK. [2]Council for Scientific and Industrial Research - Oil Palm Research Institute (CSIR-OPRI), Sekondi, Ghana. [3]CABI, Nosworthy Way, Wallingford, Oxfordshire, OX10 8DE, UK.

Accepted 10 October, 2013

Cape St. Paul Wilt Disease (CSPWD) is a major factor that impacts on coconut productivity in Ghana. Novel markers that might be specific for a promising variety of coconut or that could validate the efficacy of crosses would be valuable for confidence in the identity of palms. This study couples the discovery of such a marker with a high throughput genotyping system based on high resolution melt curve analysis. Using oligonucleotides designed against kinase subdomains of receptor like kinases (RLKs) of other plant species, eight putative RLK genes were isolated from coconut, and the intron sequence of one of these analysed in more detail. Three single nucleotide polymorphisms (SNPs) were identified within this intron that could be used as a tractable marker to differentiate two distinct genotypes, and which could be differentiated using high resolution melt curve analysis. Analysis of different varieties of coconut used in the breeding programme included promising hybrids such as Sri Lanka Green Dwarf x Vanuatu Tall. F1 crosses between these palms had been self pollinated to generate F2 populations. Genotyping of palms at the RLK marker suggested that some F2 offspring of parent F1 palms may have been sired via cross pollination from neighbouring palms, a possibility that would bear significance for such breeding programmes.

Key words: Coconut, genetic markers, receptor-like kinases, high resolution melt curve analysis, single nucleotide polymorphisms (SNPs).

INTRODUCTION

Coconut (*Cocos nucifera* L.; Arecaeae) is an important crop in coastal tropical areas where it supports the livelihoods of many poor people and helps sustain the environment. It can be grown (with minimal capital outlay) in poor soils where no other crops would survive, and is a source of material for food, drink and shelter, providing essential nutrients and also potential income. Twelve million hectares of coconut are grown worldwide and 96% of the farmers are smallholders, tending less than four hectares (Eden-Green, 1999; Dery et al., 2005). A major factor that impacts on coconut productivity in Ghana and throughout Africa, as well as in the Caribbean and

*Corresponding author. E-mail: matthew.dickinson@nottingham.ac.uk

Abbreviations: LYD, Lethal yellowing disease; **CSPWD,** Cape St Paul wilt disease; **LY,** lethal yellowing; **MYD,** Malayan dwarf; **PNT,** Panama tall; **VTT,** Vanuatu tall; **SGD,** Sri Lanka Green Dwarf; **SSRs,** small sequence repeats; **WAT,** West Africa tall; **CTAB,** cetyl trimethyl ammonium bromide; **BLAST,** Basic local alignment search tool; **PCR,** polymerase chain reaction; **NCBI,** National Center for Biotechnology Information; **ESTs**, expressed sequence tags

Central America, is disease and in particular the Lethal Yellowing (LY) Like diseases caused by phytoplasmas. The symptoms of the disease are characterised by premature fruit drop and blackening of new inflore-scences followed by yellowing of the leaves until the crown dies to result in bare trucks. Similar diseases in Africa are referred to as Lethal yellowing like because the phytoplasmas involved are different strains. In Ghana, the disease was first noted in 1932 and is referred to locally as Cape St Paul wilt disease (CSPWD). The disease has since spread westward and the disease front is now close to the Côte d'Ivoire border. Millions of trees have died and LY / LYDs are regarded as the most significant factors impacting on coconut production worldwide (Oropeza et al., 2005). Lethal Yellowing Disease (LYD) is currently going through a second epiphytotic in the Caribbean, in particular in Jamaica, where two thirds of palms (over three million) have been killed because the resistance in the Malayan yellow dwarf (MYD) and MayPan hybrids that were introduced to combat the disease following the first epiphytotic in the early 1970s has broken down. Past breakdown of resistance in coconut breeding programmes against simi-lar diseases have been attributed to pathogen evolution (Broschat et al., 2002), but could also occur through impurities or errors in breeding strategies. In Jamaica, heavy losses of MayPan hybrid palms occurred due to Lethal Yellowing despite them being planted extensively to control spread of the disease. Their failure has been partly explained by the genetic contamination of the Panama tall (PNT), the pollen parent, with pollen from the susceptible Jamaican Tall ecotype and a large per-centage of off-types observed in the MYD mother palms (Broschat et al., 2002; Baudouin et al., 2008; Lebrun et al., 2008).

Replanting with resistant palms has proved to be the most effective means to deal with LY/LYD, and certainly in Ghana farmers have been prepared to take the risk of replanting with varieties that are not guaranteed, such is the demand for coconut (Eden-Green, 1999). In the Caribbean, the use of MYD and MYD hybrids (MayPans), had been largely effective until the recent epiphytotic. In Africa, where different strains of phytoplasma occur, different sources of resistance are required. In Ghana's coconut breeding programme, screening and selection for resistance to CSPWD has led to the identification of two promising hybrids. The MYDx Vanuatu tall (VTT) hybrid was thought to be promising for resistance to CSPWD disease, but resistance seems to have been either broken down or compromised. The Sri Lanka Green Dwarf (SGD) xVTT hybrid is now believed to be more promising (Dery et al., 2008), and a population of SGDxVTT F2s has been developed and is being screened for resistance to the disease.

In Ghana, the materials used in the breeding programs have not been characterised and it is expedient to develop a reliable set of molecular markers and techni-ques that can be used to discriminate between the different varieties. In a disease control context, it is essential that hybrids or supposed resistant/tolerant materials distri-buted to farmers are exactly what they were claimed to be and this requires characterisation of the breeding materials. Thus, the need to ensure genetic purity in breeding programs is important and this requires having the appropriate tools such as molecular markers for cultivar identification. Use of conserved regions to clone novel genes is a valuable strategy for enabling the dis-covery of novel markers, especially in species with limited molecular resources available. Microsatellites, or small sequence repeats (SSRs) have been used for evaluating population diversity in several continents (Gunn et al., 2011; Liu et al., 2011; Ribeiro et al., 2010).

However, in this work, we attempted to develop an alternative strategy using single nucleotide polymor-phisms in specific genes. The conservation of amino acid sequences has enabled the isolation of resistance gene analogues from different species with degenerate primers (Collins et al., 1998; Shen et al., 1998), and these RLK genes were selected for this study because of their known involvement in defence responses to other types of plant pathogens, and potential involvement in defence against phytoplasmas. Similar approaches, such as the use of SNPs based on WRKY gene sequences have previously been developed in cocoa and coconut (Borrone et al., 2004; Mauro-Herrera et al., 2006). Although SNPs are the most abundant markers, the lack of mass coconut genome sequences has hampered their discovery.

The objectives of this study were to couple the discovery of such a marker with a high throughput genotyping system based on high resolution melt curve analysis. This is a powerful tool for detection of polymer-phisms, including those that are subtle from small differences in size or single nucleotide polymorphisms. The aim is to find novel markers that might be specific for a variety of coconut, particularly those used in the coconut breeding programme in Ghana.

MATERIALS AND METHODS

Coconut varieties and sampling of plant material

Palms being used in the coconut breeding programme in Ghana were used in this study, at sites in the Western Region of Ghana. The palms included samples of West Africa tall (WAT; 100 palms from site Fasin), VTT from field site Agona Junction, MYD from site Aiyinase, SGD from Bamiankor and Bonsaso, F1 palms of the cross MYDxVTT (from sites Daboase and Agona Junction), F1 palms of SGDxVTT (from an experimental plot at Agona Junction), and F2 palms of selected SGDxVTT F1 palms (Agona Junction) that were self pollinated in 2008 (SGDxVTT F2 palms at an experimental plot at site Asebu, an area known to be a focus for CSPWD). Coconut trunk tissues were collected following the method of Nipah et al. (2007) with slight modifications as follows: A motorised drill fitted with a sterilised drill bit was used to bore a hole of about 10 cm into the trunk of the coconut at a height of about 1 m above ground level; in this process the phloem tissues are chipped out in the form of sawdust. To prevent cross contamination from palm to palm, the

Table 1. Oligonucleotides used in this study for the discovery of RLK sequences or introns downstream of these sequences, amplifying one intron specifically, identifying which RLK sequence this intron was downstream of, and amplification of a small portion of this intron for the purpose of diagnosing the genotype of three SNPs therein.

Oligonucleotide	Sequence (5'-3')	Source	Purpose
RLKF	ATCGGKAARGGCGGMGCKGGRATYGTSTAC	a	Obtaining RLK coding regions
RLKR	GGSGCGATGTAKCCRTARGAGCCAGC	a	
RLKR4c	GCTGGCTCYTAYGGMTACATCGCSCC	Novel	Obtaining introns downstream of coding regions
KIXR1	AARCTRTASACRTCRCTYTTCTCRTC	Novel	
RLKF	ATCGGKAARGGCGGMGCKGGRATYGTSTAC	a	Identifying the RLK coding region the intron lies downstream of coding regions
KIXR1	AARCTRTASACRTCRCTYTTCTCRTC	Novel	
CnRLKintF1	GGTTGTTATTTGGGATTCAAC	Novel	Amplifying most of intron 1 for comparison of varieties
CnRLKintR1	AACGGAAGAGAATAAATTATGACA	Novel	
CnRLKintF2	CCAACTTTAGCTTATTTGTCAAAC	Novel	Melt curve analysis of SNPs in intron
CnRLKintR2	TCTTCTCGTCCACCTTCAG	Novel	

Oligonucleotides denoted 'a' are sourced from Yamamoto and Knap (2001).

drill bits were washed in water, rinsed in 0.5% sodium hypochlorite and flamed to red hot before cooling in ethanol.

DNA extraction

To obtain genomic DNA from the coconut tissues, a 2 mL tube (Starlab, USA) containing about 6-8 glass beads (Sigma, USA) was half filled with coconut tissue and ground in a fastprep^{TM} machine (Thermo electron corporation, Massachusetts, USA) at 6500 rpm for 45 s. Cetyl trimethyl ammonium bromide (CTAB) buffer (700 µL; Doyle and Doyle, 1990) was added to the tissues and homogenised again at 6500 rpm for 45 s. DNA was then extracted using chloroform/iso-amyl alcohol and precipitated with isopropyl alcohol using a protocol of Daire et al. (1997).

Amplification of putative RLKs by polymerase chain reaction (PCR)

For discovery of putative RLK genes from coconut, the oligonucleotides RLKF and RLKR were used (Table 1; Yamamoto and Knap, 2001). DNA from MYD (as an initial reference sample) was amplified as follows. Reactions (25 µl volume) comprised GoTaq reaction buffer, $MgCl_2$ (at a final concentration of 1.5 mM), dNTPs (0.2 mM each), 0.625 units GoTaq DNA Polymerase, oligonucleotides (0.2 µM each), 0.5 µl template DNA. Reactions were carried out using a Flexigene Thermal Cycler (Techne) with an initial denaturation temperature of 94°C for 3 min, followed by 35 cycles of 94°C for 60 s, 50°C for 60 s and 72°C for 3 min, and a final extension of 72°C for 10 min. Products were analysed by agarose gel electrophoresis (1.5% w/v, as compared to Hyperladder II; Starlab, UK). DNA from SGD, VTT and WAT was also used to obtain further sequences. To obtain sequences downstream of oligonucleotide RLKR, the reverse complement of this sequence was used as a new forward oligonucleotide (RLKRc; Table 1) in combination with a novel reverse oligonucleotide (KIXR1; Table 1), which was designed from alignments of the coding sequences of top matches from the Basic Local Alignment Search Tool (BLAST) searching of putative CnRLK sequences (including soybean, vine, rubber and White Campion), with RLK sequences from Arabidopsis thaliana, rice, bamboo and soybean. DNA from MYD was initially amplified using the standard amplification conditions and analysed by agarose gel electrophoresis as described above. Candidate amplicons were PCR purified (using a PCR purification kit, following the manufacturers instructions; Sigma, Poole, UK) or were gel purified (using a gel purification kit; Sigma). Amplicons were cloned and sequenced as described below.

In order to compare the sequence of a putative intron between coconut varieties, oligonucleotides were designed covering most of the sequence (excluding the conserved regions around RLKRc and KIXR1; using oligonucleotides CnRLKintF1 and CnRLKintR1; Table 1). PCR was carried out following the procedure described above, followed by PCR purification and sequencing, or cloning then sequencing. To find out which putative RLK candidate an intron sequence was downstream of, DNA was also amplified with oligonucleotides RLKF and KIXR1, and cloned sequences of the resulting product (which were greater than 1kb) were compared with sequences of putative RLKs and putative introns. Cloning and sequencing of amplicons of this size is less efficient and reliable than smaller fragments (respectively), which is why amplification was done using RLKRc and KIXR1 in the first instance.

Cloning and sequencing of PCR products

Purified fragments were cloned using the pGEM-T easy vector system (Promega), according to the manufacturer's instructions, transformed into *Escherichia coli* competent cells JM109 (Promega Corp.), cultured, and subjected to blue white selection and reculturing. Cloned fragments were then directly reamplified from *E. coli* colonies by PCR using oligonucleotides M13F (5'-GTAAAACGACGGCCAGT-3') and M13R (5'-CAGGAAACAGCTATGAC-3'), and assessed by agarose gel electrophoresis to confirm that ligation had been successful. Samples were purified using a PCR purification kit (Sigma), and sequenced by MWG Eurofins using primers M13F and/or M13R. Sequence alignments were made using Clustalw alignment in Bioedit 7.1.3.0 (Hall, 1999). Sequencing output was also assessed manually by examining SCF fluorescence traces, in order to confirm that sequencing data were of good quality and that bases were unambiguous. This was especially important where a candidate SNP was thought to occur, since in the case where a sample may be heterozygous for a SNP, fluorescence traces could overlap at the base position and base calling alone could not be relied upon.

Sequence analysis

Comparison of the sequences of putative RLKs was carried out by

searching for homology with known sequences of RLKs from other species by performing blastx (translated nucleotide) searches of the National Center for Biotechnology Information (NCBI) protein databases. Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 5 (Tamura et al., 2011). Phylogeny reconstruction was done with neighbour joining using the 'maximum composite likelihood' stochastic model, and a phylogenetic tree constructed. Sequences were also compared to the rice genome using the Rice Genome Annotation Project database (http://rapdblegacy.dna.affrc.go.jp/tools/blast/) in order to determine whether it may be likely that genomic sequences may contain non-coding regions.

High resolution melt curve analysis

For accurate melt curve analysis of products, a small amplicon is required, so oligonucleotides were designed within the putative intron of CnRLK1 flanking the location of putative SNPs (CnRLKintF2; Table 1, CnRLKintR2; Table 1). CnRLKintF2 was designed within the intron to provide specificity, with CnRLKintR2 straddling coding and non-coding sequence (the priority was that the predicted reaction would be efficient, and amplicon small, and these criteria resulted in a reverse oligonucleotide in a conserved region). A PCR product of 227 base pairs in size was predicted and verified using standard PCR conditions and agarose gel electrophoresis described above. Subsequent reactions were carried out on 96 well opaque reaction plates (BioFire Diagnostics, Inc., formerly Idaho Technology, Utah, USA) with a Flexigene Thermal Cycler (Techne), then melt analysis carried out using a LightScanner (BioFire Diagnostics). Reactions (10 µl volume) comprised 5 µl SensiMix HRM buffer (Bioline), 0.5 µl EvaGreen dye, 0.5 µl template DNA, oligonucleotides to a final concentration of 0.5 µM each. Each reaction was overlaid with 15 µl mineral oil, an adhesive optical lid secured and the plate centrifuged briefly. The following PCR conditions were used: 94°C for 10 min, followed by 40 cycles of 94°C for 15 s, 58°C for 10 s and 72°C for 10 s, and a final step of 25°C for 30 s. Following amplification, a melt curve was immediately performed using a LightScanner (BioFire Diagnostics) to produce the melting profiles of amplicons by detecting fluorescence therein, measured in real-time during a programme ramping the temperature from 60 to 90°C at the default Lightscanner melting rate ($0.1^{o}C\ s^{-1}$). Fluorescence data were analysed using LightScanner Software according to the manufacturers' instructions, to obtain melt curves that were normalised for comparison to those generated using reference samples of known genotype. Differences between alleles could easily be discriminated by eye (Figure 2c) and were grouped manually for export.

Mapping of palm locations

To accurately represent the location of palms at field site Agona, GPS mapping of the site was conducted. Garmin eTrex hand held GPS units were used to record the perimeters of plots at field site Agona, and record the locations of certain palms. Mapping data were analysed for presentation using GPSMapEdit.

RESULTS

Discovery and characterization of RLK sequences in coconut

With the aim of identifying SNPs or similar markers in genes of coconut varieties, putative receptor-like kinases were obtained. Oligonucleotides RLKF/RLKR (specific for kinase subdomains of an RLK in soybean; Figure 1a) were used to obtain putative RLK sequences from DNA extracted from four varieties of coconut (MYD, SGD, VTT and WAT). An array of clones was obtained, mostly of the expected size of roughly 500-600 bases. From the cloned sequences analysed, sequences for eight putative unique RLKs were obtained that, upon translated BLAST searching of protein sequences in NCBI, were found to have highly significant homology with RLK protein sequences from species including tomato, vine, *Arabidopsis thaliana*, maize and castor oil plant, *Ricinus communis* (Table 2, Figure 1b). Comparisons between nucleotide sequences of each putative RLK obtained independently from each variety failed to reveal differences between them; there were no differences in the sequences of each putative RLK obtained from cloning of sequences from each variety.

Comparison of the sequence of each putative RLK with the rice genome revealed that five had the highest homology to rice kinases that contained large introns immediately downstream of the region of homology. To determine whether an intron might be present down-stream of putative RLK genes in coconut, a novel oligonucleotide was designed within the region of kinase subdomain KIX (downstream of the obtained putative RLKs sequences; 'CnKIXR1'), and this was used together with an oligonucelotide in kinase subdomain VII ('CnRLKR4c', the reverse complement of RLKR; Figure 2a) so that if an intron was present in the nucleotide sequences between these two kinase domains, it would be apparent by a large amplicon.

Using these oligonucleotides, DNA was amplified from a reference sample of MYD, generating three amplicons of approximately 400 to 650 bases, which were cloned and sequenced. Apart from the regions containing the oligonucleotides (kinase subdomains VIII and KIX), there was found to be no homology between sequences, and BLAST searching revealed no significant similarities, which suggested they might be non-coding DNA. One sequence was selected for further analysis and oligonucleotides RLKF and KIXR1 were used to verify that the sequence discovered was from within a coconut RLK gene, and determine which putative RLK lied upstream of it. This was found to be immediately downstream of the sequence obtained for CnRLK1, and therefore appeared to be an intron in gene CnRLK1, of 530 bases in size.

Discovery and analysis of a SNP marker in reference palm samples

Comparison of sequences of the apparent intron in CnRLK1 (by sequencing PCR products of primers CnRLKintF1 and CnRLKintR1) from up to eight samples of each variety gave a completely conserved sequence in samples of SGD, MYD and VTT. However, alignments

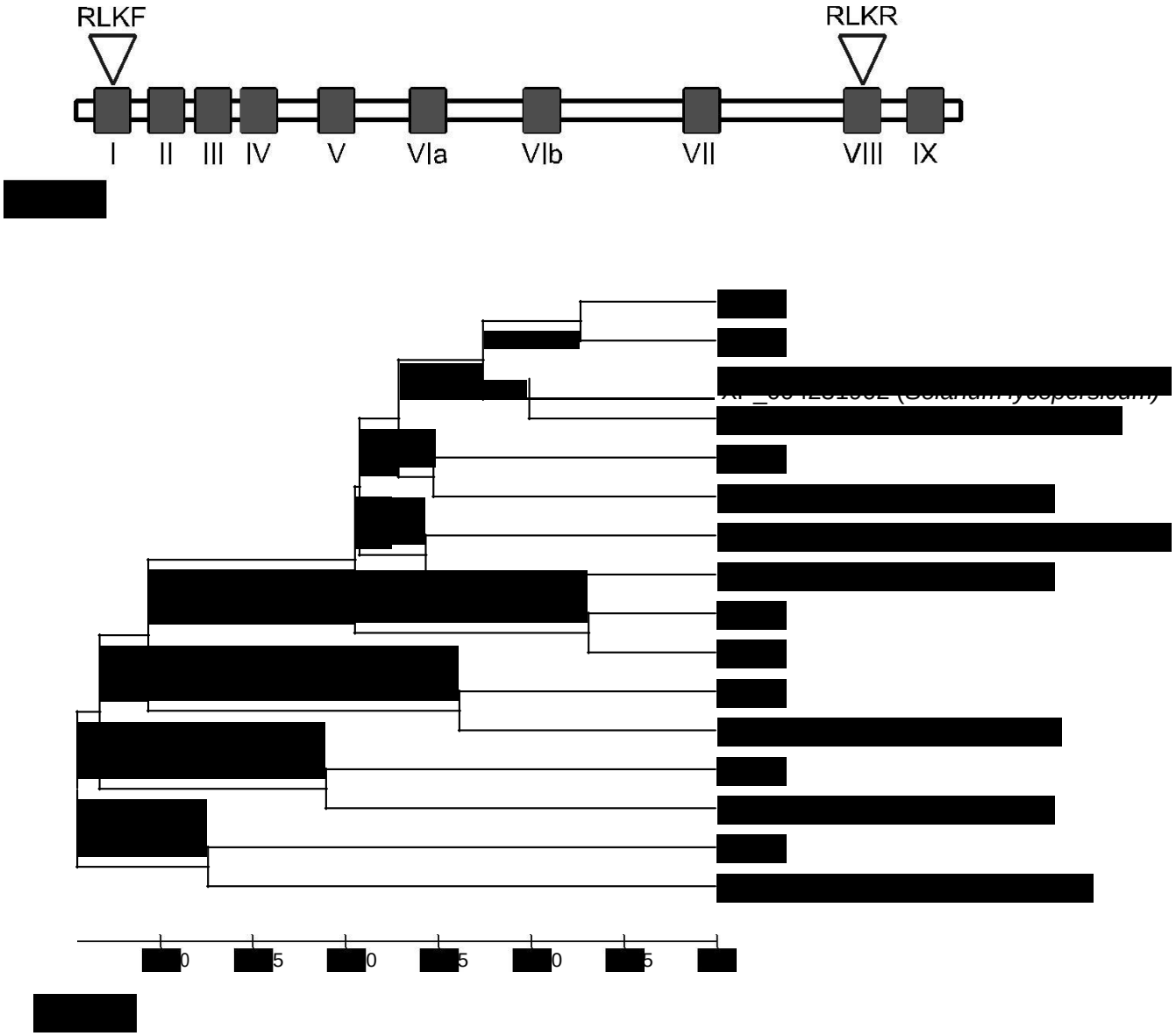

Figure 1. (a) Representation of the protein structure of RLKs from Arabidopsis and soybean, showing the positions of primers RLKF and RLKR, used to isolate coding regions of RLK genes from coconut (after Yamamoto and Knap, 2001, figure not to scale, codons for approximately 200 amino acids lie between the positions of RLKF and RLKR). **(b)** Phylogenetic analysis of the nucleotide sequences of eight unique putative RLK genes isolated from coconut, along with nucleotide sequences with greatest homology to each, from blastx analysis (see Table 2 for details).

showed that sequences from several WAT palms differed from these at three positions: with T, A and A present in positions 460, 491 and 516 in WAT, as opposed to C, G and G in other varieties (Figure 2b). Analysis of sequencing trace data showed that one of the eight WAT samples was from a palm that possessed both the TAA and CGG genotypes, suggestive of a palm that was heterozygous for all three SNPs. This was evident in sequencing trace data, which showed overlapping fluorescence peaks at each SNP (and these peaks were of half the height of surrounding peaks). Cloning of the putative intron in this sample (and two others subsequently identified via melt curve analysis, as described below) showed that sequences cloned from an apparently heterozygous sample were always present as CGG or TAA, not a mixture of polymorphisms at the SNPs, showing that only three genotypes had been found: TAA/TAA, CGG/CGG and TAA/CGG.

A high resolution melt curve analysis method was used to genotype more samples from each site at this marker, with greater throughput than by sequencing. Oligonucleotides were designed that would be specific for the in-

Table 2. BLAST analysis of top protein matches of each of eight unique sequences obtained using oligonucleotides RLKF and RLKR.

Putative RLK*	Top BLAST hit: Accession	Description	Max score	Total score	Query coverage	E value	Max ident
CnRLK1	XP_004231962	Leucine rich repeat receptor-like serine/threonine-protein kinase BAM1-like [*Solanum lycopersicum*]	347	347	99%	3e-110	95%
CnRLK2	XP_002531999	Receptor protein kinase CLAVATA1 precursor, putative [*Ricinus communis*]	338	338	99%	9e-107	95%
CnRLK3	XP_002264952	PREDICTED: Leucine-rich repeat receptor-like serine/threonine-protein kinase BAM3 [*Vitis vinifera*]	313	313	99%	1e-97	87%
CnRLK4	XP_004238370	PREDICTED: receptor protein kinase CLAVATA1-like [*Solanum lycopersicum*]	298	298	100%	3e-92	84%
CnRLK5	XP_002279563	PREDICTED: Receptor protein kinase CLAVATA1-like [*Vitis vinifera*]	308	308	99%	8e-96	87%
CnRLK6	XP_003632392	PREDICTED: leucine-rich repeat receptor-like protein kinase PXL2-like [*Vitis vinifera*]	271	271	99%	7e-82	75%
CnRLK7	BAD94141	Leucine-rich repeat receptor-like kinase At1g09970 [*Arabidopsis thaliana*]	244	244	99%	8e-77	64%
CnRLK8	AFW57819**	Putative leucine-rich repeat receptor protein kinase family protein [*Zea mays*]	306	306	99%	5e-94	82%

Top hit is shown for each putative CnRLK1-8, from translated BLAST (blastx) analysis (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=TranslationsandPROGRAM=blastxandBLAST_PROGRAMS=blastxandPAGE_TYPE=BlastSearchandSHOW_DEFAULTS=onandBLAST_SPEC=). *NCBI nucleotide accession numbers are sequentially numbered as follows: CnRLK1-8: KC020611-KC020618. ** Maize Genome Database transcript name GRMZM2G034155_T01: http://maizegdb.org/cgi-bin/displaygenemodelrecord.cgi?id=GRMZM2G034155_P01.

tron, but would amplify a short sequence flanking the three SNPs. With the incorporation of a fluorescent intercalating probe into amplicons, their melting profile was measured in real time following amplification, with higher melting temperature (tm) predicted for samples of CGG, and lower tm expected in samples of TAA. High resolution melting distinguished between genotypes and was found to consistently show three melt curve patterns that were easily distinguishable (Figure 2c). Reactions performed on DNA from palms of known genotype CGG/CGG were found to have the melting profile showing the highest melt temperature. Palms of TAA/TAA genotype clearly showed the lower melting temperature, and reactions from palms shown by sequencing to be heterozygous for TAA and CGG had melt curve profiles intermediate between the other two genotypes. Comparisons of melt curve results and sequencing data, and addition of more samples of each variety, concurred with the initial findings that all tested MYD, SGD and VTT palms were of genotype CGG/CGG (Table 3). Of the 100 WAT palms tested, 97 were TAA and three were heterozygous.

Analysis of hybrid palms at experimental plots

Since the triple-SNP marker discovered in the putative intron of CnRLK1 could discriminate between WAT and other palms, it was decided to test its use for characterising supposed hybrids that were being used on a smallholder farm (MYDxVTT at Daboase) and at experimental trial sites (SGDxVTT at Agona, SGDxVTT F2 palms at Asebu). Using melt curve analyses, genotyping of MYDxVTT palms at Daboase showed that of 82 palms sampled, 50 were CGG/CGG and 32 were TAA/CGG. Melt curve analysis of samples from Agona showed that 47 SGDxVTTs were all of genotype CGG/CGG (Table 3). Another large plot of palms at Agona contains MYDxVTT palms (Figure 3), sixteen of which were genotyped (with 10 being CGG and 6 heterozygous; TAA/CGG; Table 3). The sixteen MYDxVTT palms were taken from the southern edge of the plot, bordering the SGDxVTT plot (Figure 3). Among other palms at the site were three WAT palms of about 30 years old, approximately 20 metres south west of the SGDxVTT plot (Figure 3). These palms were uphill of the plot of SGDxVTTs and separated from them by some mixed vegetation of up to 2.5 metres height. The three WAT palms were found to carry the TAA/TAA genotype.

In 2008, 14 of the SGDxVTT F1 palms at Agona were chosen to be self pollinated for the generation of a population of F2 SGDxVTT palms as part of the coconut breeding programme. The location of these parent palms is shown in Figure 3. The resulting F2 progeny were planted in Asebu later in 2008. In total, approximately 230 palms were planted (on average, 14 F2 palms were obtained

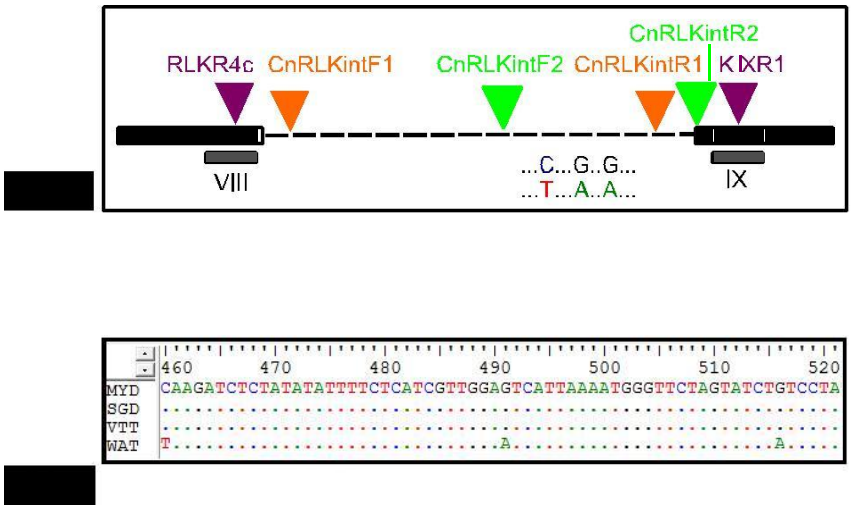

Figure 2. (a). Position of olicognucleotides used to amplify and sequence the intron from CnRLK1 (RLKR4c and KIXR1; coloured purple), and position of oligonucleotides used for specifically amplifying this intron, for comparison between varieties (CnRLKintF1, CnRLKintR1, coloured orange) oligonucleotides used for HRM analyses (CnRLKintF2, CnRLKintR2, coloured green). Black solid bars represent coding regions, dashed line shows intron position and grey boxes indicate the position of kinase subdomains VIII and IX. The approximate position of three SNPs is also indicated. Figure not to scale. (b). Section of sequences of intron CnRLK1 in representative samples of palms of MYD, SGD, VTT and WAT. The position of three SNPs is visible at bases 460, 491 and 516. Sequences containing the intron and partial CDS are presented in GenBank: NCBI nucleotide accession numbers KC020619 and KC020620. (c). Shifted melt curve plots of representative samples from each genotype sampled. Duplicate reactions of 27 extracts are shown, including MYD, SGD, VTT, WAT. TAA/TAA is displayed as red plots, showing the lowest melting temperatures, CGG/CGG is shown as the blue plots (highest melting temperatures) and grey plots show the intermediate melting temperatures of TAA/CGG heterozygous samples.

from each parent palm). Of the 209 palms surviving to produce at least two leaves, sampling and genotyping at the marker in CnRLK1 via melt curve analysis showed that 202 carried the CGG/CGG allele, whereas 7 were found to be heterozygous. The SGDxVTT F2 palms that were found to be heterozygous were sired by SGDxVTT

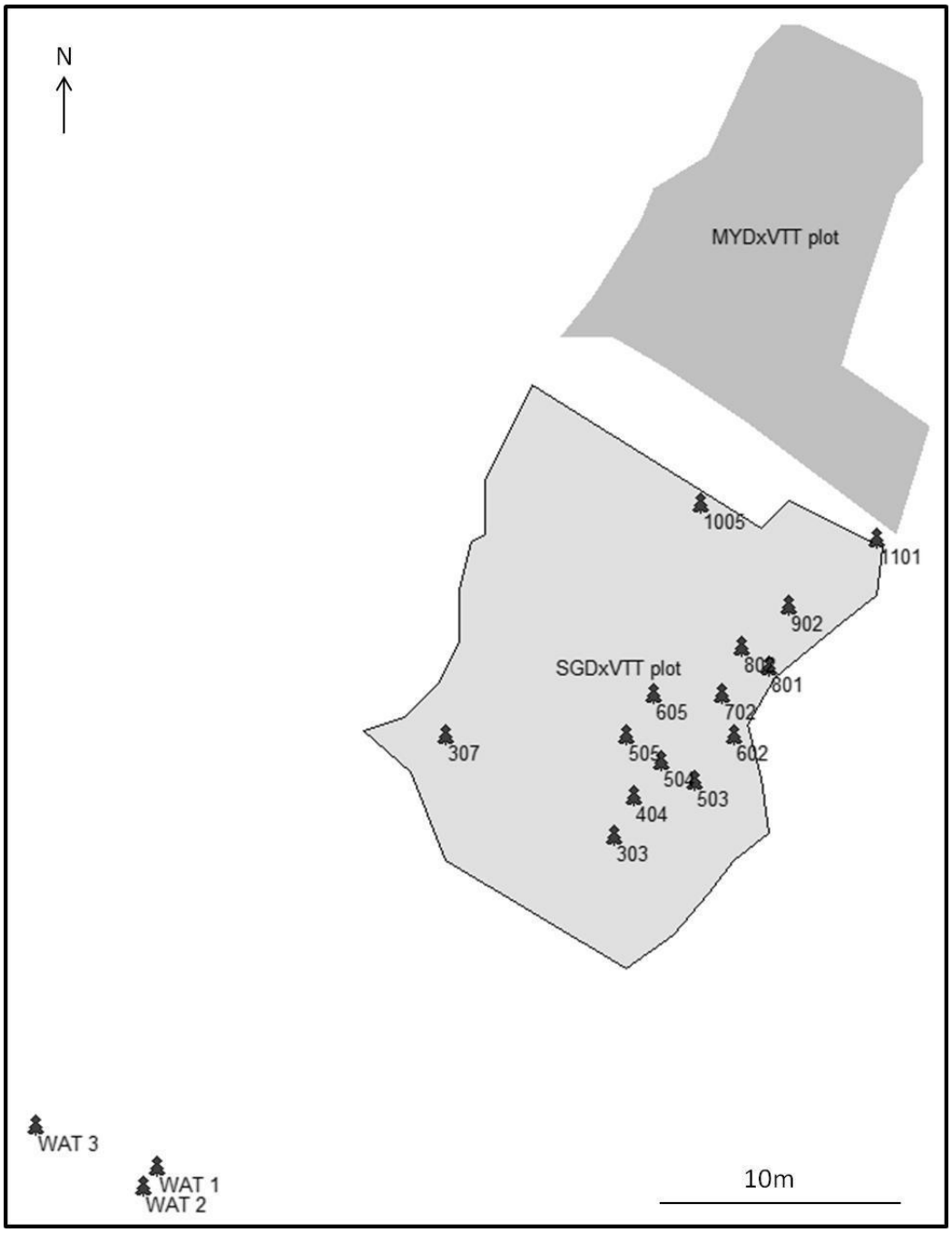

Figure 3. Position of palm plots and experimental palms at field site 'Agona', mapped by GPS. Plots of SGDxVTT and MYDxVTT F1 palms are shown, with the position of those SGDxVTT palms used to generate populations of F2 palms indicated (palms 303-1101). The positions of three WAT palms at this site are also shown. Figure is to scale.

(F1, Agona) palms 303, 307, 702, and 802.

DISCUSSION

This study aimed to find novel markers that might be specific for a variety of coconut, ideal for the development of markers that can allow for the validation of the identity of palms of promising characteristics in their resistance to Cape St. Paul Wilt Disease. Several genes were investigated in order to discover novel markers, following on from analyses of WRKY, ribosomal or plastid genes examined in similar work in coconut and other species (Demesure et al., 1995; Mauro-Herrera et al., 2006;

Table 3. Genotypes of all sampled coconut palms at the marker in the intron of CnRLK1. Genotyping was carried out using high resolution melt curve analysis (oligonucleotides CnRLKintF2/CnRLKintR2).

Variety	n	CGG/CGG	TAA/CGG	TAA/TAA
Reference samples				
MYD	8	8		
SGD	20	20		
VTT*	8	8		
WAT	100		3	97
Daboase				
MYDxVTT	82	50	32	
Agona				
SGDxVTT	47	47		
MYDxVTT	16	10	6	
WAT	3			3
Asebu				
SGDxVTT (F2s)	209	202	7	

* VTT palms also from Agona.

Calvino and Downie, 2007; Meerow et al., 2009). Using oligonucleotides designed against conserved domains of RLKs in other species (Yamamoto and Knap, 2001), this study identified eight putative RLKs in coconut. Whilst the coding sequences identified did not differ between variety, several of these genes were predicted to contain at least one large intron, one of which was analysed in detail and found to contain three SNPs that were present as two different alleles, or in a heterozygous combination. That only two combinations of bases were found at the three SNPs (there were clearly two variants in the population tested since the three SNPs appear as triplets, not a combination thereof) indicates that two versions of the CnRLK1 gene exist in the palms sampled. One allele was found to be characteristic of (though not exclusive to) the Ghanaian West African Tall samples, which are susceptible to Cape St. Paul Wilt disease (CSPWD). Although it is possible that the coconut varieties tested are too heterogeneous to permit discovery of a fully variety-specific marker (as appears likely from the variability visible in studies working with SSR markers in coconut), a simple, tractable maker such as the one described here can be used to great effect in verifying the pedigree of a palm, as compared to its parents.

Introns are excellent candidates for discovery of polymorphisms because they would not cause changes at protein level and so are subject to faster mutation rates. The potential for introns has been studied in intron flanking expressed sequence tags (ESTs) in rice (Tamura et al., 2009), as has the potential for intron length polymorphisms (for example, ILP in resistance gene analogues in wheat; Shang et al., 2010). Use of melt curve analysis in genotyping these markers is becoming increasingly acknowledged for its potential.

The discovery of individuals among the F2 population of SGDxVTT palms that were heterozygous for the marker in the intron of CnRLK1, despite all fourteen parent palms being homozygous for the CGG allele, suggests that cross pollination of some of the parent palms may have occurred. Tall varieties are primarily outbreeding via wind pollination (or potentially pollinated via insects) whereas dwarf varieties of coconut are mostly self-pollinating (Baudouin and Santos, 2005; Ramanatha Rao, 2005). There is a further possibility that those F2 palms found to be CGG/CGG could also have cross pollinated with neighbouring SGDxVTT or MYDxVTT palms at Agona. The discovery of three WAT palms at the Agona site was a further cause for concern, since they were likely to be around thirty years old and may have been releasing pollen when the SGDxVTT parent palms were selected for self pollination. The prevailing winds in the coastal region of Ghana are from the south west making the prospect for pollen spread across the plot of SGDxVTTs feasible in this wind pollinated species. MYDxVTTs were often found to be heterozygous, so could also be a source for the possible contamination detected at Agona. Development and analysis of further markers will establish the extent to which this may have occurred.

The possibility of cross pollination could have implications for the coconut breeding programme since it suggests that at least some palms of the promising SGDxVTT hybrid used for generation of material for screening for resistance to CSPWD may have been com-

promised, and possibly from susceptible palms. However, it should be noted that careful precautions are taken at sites such as these to prevent cross pollination; normally, to prevent extraneous pollen from contaminating the crosses, all palms within 300 metres are castrated since under natural conditions coconut pollen can travel over 300 metres (Mantriratne, 1965; de Nucé de Lamothe and Rognon, 1975; Ramanatha Rao, 2005). Furthermore, while assisted pollination (without precautionary bagging) can result in pollen contamination (unless very large plot sizes are used) (de Nucé de Lamothe and Rognon, 1975; Ramanatha Rao, 2005) the crosses carried out are followed by bagging of tassels. Therefore the occurrence of cross pollination is not confirmed and requires more attention: similar markers will be identified to further investigate this. It is also possible that the purity of the MYDxVTT hybrid palms at Daboase, which were deployed in a farm in an area of disease pressure, may also have been compromised. For example, planting material may have been supplemented with self-seed from local WATs, and it is these WATs rather than the MYDxVTT that have succumbed to the CSPWD. This is also under examination.

The mechanism with which to identify markers that might be used to validate the pedigree of a palm is, at present, valid and valuable in a situation where plant or DNA material from both parent and offspring is available. The collection of eight RLK sequences obtained during this study will be useful for further marker discovery and we are assessing these and introns therein with the aim of identifying further useful markers for the coconut breeding programme, using such high throughput genotyping methods as those described here. Simple genotyping strategies will be valuable in breeding still, especially transferable technologies: melt reactions similar to those carried out in this work can be done on increasing number of machines, such as portable amplification machines including isothermal amplification (for example, using the Optigene Genie II machine). Such work could be carried out in-country on portable, affordable machines, and we are assessing the feasibility of such an application.

Ultimately it is hoped that, since these palms are planted in the diseased area of Asebu, markers associated with resistance or tolerance can be identified based on a marriage of the genotyping and diagnostics being developed (Yankey et al., 2011). The advance of sequencing technologies will be used to facilitate the discovery of useful novel sequences, but for subsistence crops (as compared with cash crops) amenable strategies such as the one described here offer a valuable resource. The work here has shown a strategy for the discovery of SNPs in introns, coupled with a high throughput analysis method that could be adapted for use in breeding programmes. In future work, it should be possible to identify SNP-based markers specifically linked to phytoplasma resistance, and to develop a strategy in which breeding materials and/or planting materials distributed to growers are tested to confirm that they are resistant. Such markers would be able to identify contaminated breeding material before it is provided to growers so that the situation with resistance breakdown in the MayPan palm can be avoided.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge funding from the Biotechnology and Biological Sciences Research Council (BBSRC) and Department for International Development (DFID) (Sustainable Agriculture Research for International Development, SARID, Initiative). We thank Sylvester Apem-Darko, Joseph Opoku and Alex Watson-Lazowski for technical assistance.

REFERENCES

Baudouin L, Lebrun P, Myrie W, Been B, Dollet M (2008). The Panama Tall and the Maypan hybrid coconut in Jamaica: did genetic contamination cause a loss of resistance to lethal yellowing? Euphytica 163:353-360.

Baudouin L, Santos G (2005). Morphometric methods of determining diversity in coconut. In Coconut Genetic Resources. Batugal P., Ramanatha Rao V.R. and Oliver J., eds., 209-224.

Borrone JW, Kuhn DN, Schnell RJ (2004). Isolation, characterization, and development of WRKY genes as useful genetic markers in Theobroma cacao. Theor. Appl. Genet. 109:495-507.

Broschat TK, Harrison NA, Donselman H (2002). Losses to lethal yellowing cast doubt on coconut cultivar resistance. Palms 46:185-189.

Calvino CI, Downie SR (2007). Circumscription and phylogeny of Apiaceae subfamily Saniculoideae based on chloroplast DNA sequences. Mol. Phylogenet. Evol. 44:175-191.

Collins NC, Webb CA, Seah S, Ellis JG, Hulbert SH, Pryor A (1998). The isolation and mapping of disease resistance gene analogs in Maize. Mol. Plant-Microbe Interact. 11:968-978.

Daire X, Clair D, Reinert W (1997). Detection and differentiation of grapevine yellows phytoplasmas belonging to the elm yellows group and to the stolbur subgroup by PCR amplification of non-ribosomal DNA. Eur. J. Plant Pathol. 103:507-514.

de Nucé de Lamothe M, Rognon F (1975). Assisted pollination and contamination by undesirable pollens. Oléagineux 80:359-364.

Demesure B, Sodzi N, Petit RJ (1995). A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. Mol. Ecol. 4:129-131.

Dery SK, Nipah JO, Ofori F (2005). Status of coconut genetic resources research in Ghana. In: Batugal P., Ramanatha Rao V. and Oliver J. eds. Coconut Genetic Resources. IPGRI-APO, Malaysia, 661-666.

Dery SK, Philippe R, Baudoin L, Quaicoe RN, Nkansah-Poku J, Owusu-Nipah J, Arthur R, Dare D, Yankey N, Dollet M (2008). Genetic diversity among coconut varieties for susceptibility to Cape St Paul wilt disease. Euphytica 164:1-11.

Doyle JJ, Doyle JL (1990). Isolation of plant DNA from fresh tissue. Focus 12:13-15.

Eden-Green SJ (1999). Coconut yellowing diseases, etiology and control, Africa. Summary report of the European Commission supported STD-3 projects (1992-1995), published by CTA.

Gunn BF, Baudouin L, Olsen KM (2011). Independent Origins of Cultivated Coconut (Cocos nucifera L.) in the Old World Tropics. PLoS ONE 6: e21143. doi:10.1371/journal.pone.0021143.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids Symp.Series 41:95-98.

Lebrun P, Baudouin L, Myrie W, Berger A, Dollet M (2008). Recent lethal yellowing outbreak: why is the Malayan Yellow Dwarf Coconut no longer resistant in Jamaica? Tree Genet. Gen. 4:125-131.

Liu X, Tang H, Li D, Hou L (2011). Genetic diversity of coconut cultivars in China by microsatellite (SSR) markers. Mol. Plant Breed. 2:83-91.

Mantriratne MAPP (1965). Coconut pollen. Ceylon Coconut Quarterly 16:102-110.

Mauro-Herrera M, Meerow AW, Borrone JW, Kuhn DN, Schnell RJ (2006). Ten informative markers developed from WRKY sequences in coconut (Cocos nucifera). Mol. Ecol. Notes 6:904-906.

Meerow AW, Noblick L, Borrone JW, Couvreur TLP, Mauro-Herrera M, Hahn WJ, Kuhn DN, Nakamura K, Oleas NH, Schnell RJ (2009). Phylogenetic analysis of seven WRKY genes across the palm subtribe attaleinae (Arecaceae) identifies syagrus as sister group of the coconut. PLoS ONE 4: e7353. doi:10.1371/journal.pone.0007353

Nipah JO, Jones P, Dickinson MJ (2007). Detection of lethal yellowing phytoplasma in embryos from coconut palms infected with Cape St Paul wilt disease in Ghana. Plant Pathol**.** 56:777-784.

Oropeza C, Escamilla JA, Mora G, Zizumbo D, Harrison NA (2005). Coconut lethal yellowing. 349-363. IPGRI-APO, Malaysia. pp. 349-363.

Ramanatha Rao V (2005). Coconut field genebank. In: Batugal P., Ramanatha Rao V. and Oliver J. eds. Coconut Genetic Resources. IPGRI-APO, Malaysia, 661-666.

Ribeiro FE, Baudouin L, Lebrun P, Chaves LJ, Brondani C, Zucchi MI, Vencovsky R (2010). Population structures of Brazilian tall coconut (Cocos nucifera L.) by microsatellite markers. Genet Mol. Biol. 33:696-702.

Shang W, Zhou R, Jia J, Gao L (2010). RGA-ILP, a new type of functional molecular markers in bread wheat. Euphytica 172:263-273.

Shen KA, Meyers BC, Islam-Faridi MN, Chin DB, Stelly DM, Michelmore RW (1998). Resistance gene candidates identified by PCR with degenerate oligonucleotide primers map to clusters of resistance gene in lettuce. Mol. Plant-Microbe Interact. 11:815-823.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony Methods. Mol. Biol. Evol. 28:2731-2739.

Tamura K, Yonemaru J, Hisano H, Kanamori H, King J, King IP, Tase K, Sanada Y, Komatsu T, Yamada T (2009). Development of intron-flanking EST markers for the Lolium/Festuca complex using rice genomic information. Theor. Appl. Genet. 118:1549-1560.

Yamamoto E, Knap HT (2001). Soybean receptor-like protein kinase genes: paralogous divergence of a gene family. Mol. Biol. Evol. 18:1522-1531.

Yankey EN, Swarbrick P, Dickinson M, Tomlinson J, Boonham N, Nipah JO, Quaicoe RN (2011). Improving molecular diagnostics for the detection of lethal disease phytoplasma of coconut in Ghana. Bull. Insectol. 64 (Supplement):S47-S48.