AJ ACADEMIC JOURNALS
expand your knowledge

African Journal of Biotechnology

Full Length Research Paper

Studies on some major yield responsive genes in selected rice (*Oryza* species) cultivars grown in Nigeria using candidate gene SSR-based markers approach

Zaharaddin Halilu[1*], Rabi'u Aliyu Umar[1], Riskuwa Arabu Shehu[1], Adamu Aliyu Aliero[2], Aliyu Abdullahi Turaki[3], Ahmed Hamza Balarabe[4], Abubakar Muhammad Gumi[2] and Sha'aya'u Shehu[1]

[1]Department of Biochemistry, Faculty of Science, Usmanu Dan fodiyo Universty, Sokoto. Nigeria.
[2]Department of Biological Sciences, Faculty of Science, Usmanu Dan fodiyo Universty, Sokoto Nigeria.
[3]Department of Biochemistry and Molecular Biology, Faculty of Science, Federal University of Birnin Kebbi, Kebbi Nigeria.
[4]Medway Centre for Pharmaceutical Sciences, Faculty of Engineering and Science, University of Greenwich, Medway Campus, Chatham Maritime, Kent, UK.

This study was aimed at determining the presence of five major yield responsive genes shown to be among key determinants of rice yield in ten rice cultivars grown in Nigeria using candidate gene SSRs markers (Cg-SSRs). DNAs were extracted from young leaf samples using optimized cetyltrimethylammonium bromide (CTAB)-extraction method. An established gene sequences for high yielding rice Nipponbare cultivar for the five genes were retrieved from NCBI database and used to mine for SSRS using SSR identification tool. A pair of primer were designed manually from the nucleotide sequences flanking the selected SSRS for each gene. The designed primers were used to screen for the presence of the genes. Three pair of primers (for *Gif1*, *Gn1a* and *Gw2*) were able to amplify the targeted regions. The remaining two (for *Gs3* and *Dep1*) did not amplify even the positive control. Two representative polymerase chain reaction (PCR)-products for the PCR-positive genes were sequenced and compared with established gene sequences of high yielding cultivars on NCBI database and their percentage sequence identities were determined. Phylogenetic trees (dendrograms) were generated using neighbor joining method. Overall, all the ten cultivars have both *Gif1* and *Gn1a* genes while only eight cultivars have *Gw2* gene. *Gs3* and *Dep1* need further optimization of PCR-conditions.

Key words: Polymerase chain reaction, SSRs markers, primer.

INTRODUCTION

Rice is one of the most important staple food crops that is cultivated worldwide. Rice is safe and crucial to food security for more than one-half of the world population (Miura et al., 2010). Rapid increase in population coupled

with urbanization continuously decreased our farmlands, creating ever increasing demand for rice. Thus, there is the need to improve rice varieties with high yield, high quality and stress tolerance.

Information on genetic diversity and relationships within and among crop species as well as their wild relatives is essential for the efficient utilization of plant genetic resource collections (Chan and Sun, 1997; Govindaraj et al., 2015). Breeding and genetic conservation programmes aimed at developing new improved rice cultivars can boost rice production. Grain yield is a complex trait determined by Quantitative Trait Loci (QTLS). Rice grain yield is mainly determined by three components; number of panicle per plant, grain number per panicle and grain weight (Cheung, 2014). For rice breeders, identification of suitable genotypes containing these agronomic traits for grain yield determination should be the first critical step toward breeding high yield hybrid rice.

In the last few decades, many QTLS related to rice grain yield were identified. Amongst which include *grain number 1a (gn1a)*, the first gene to be isolated that controls rice grain number reported by Ashikari et al. (2005). *Grain size 3* (*Gs3*) was the first major gene to be isolated that determine rice grain size (Fan et al., 2006), *Grain weight and width2* (*Gw2*) influences rice grain width and weight (Song et al., 2007), *Grain incomplete filling 1* (*Gif1*) regulates rice grain filling (Wang et al., 2008), *Dense and erect protein 1* (*Dep1*) determines rice panicle architecture (Huang et al., 2009; Wang et al., 2009; Zhou et al., 2009), *Osspl14* promotes panicle branching and higher grain yield in rice (Miura et al., 2010). Other genes identified include *Gs5* (Li et al., 2011), *Gs6* (Sun et al., 2013), *DST* (Li et al., 2013), *Gs2* (Hu et al., 2015), Osspl13/*Glw7* (Si et al., 2016).

Despite many policies of the Federal Government of Nigeria in the rice sector, rice production has not matched the growing domestic demand. In Nigeria, rice production has marginally increased from 5.5 million metric tons (MT) in 2015 to 5.8 million MT in 2017. The current consumption rate is 7.9 million MT (Punch Newspaper, 2017). Thus, a deficit of about 2.1 million MT is been met by importation. Nigeria spent 356 billion naira yearly on rice importation, amounting to about one billion naira per day (Vanguard Newspaper, 2013). Figures available with the Central Bank of Nigeria (CBN) showed that from January 2012 to May 2015, Nigeria had spent over 2.41 billion dollars on rice importation (Premium Times Newspaper, 2015). If Nigeria want to meet up with rice demand and save more money for other infrastructural projects. There is, therefore, the need to improve the yielding potentials of the local rice as part of the measures to increase rice production.

Understanding genetic structure and status of genetic variation of different rice varieties cultivated in Nigeria are of paramount importance as it is the first stage in a crop breeding programme. Three key elements are used in the determination of grain yield, viz: panicle size, grain number and grain size. Some QTLS have been identified as the major yield responsive genes and/or QTLS in rice plant. However, no research on these genes and/or QTLS to the best of our knowledge was conducted on rice cultivars grown in Nigeria. This informed our decision to screen for the presence of five major yield responsive QTLS in ten cultivars grown in Nigeria. These five genes are: Dense and erect protein 1 (*Dep1*), Grain number 1a (Gn 1a), Grain size 3 (*Gs3*), Grain weight and width 2 (*gw2*) and Grain incomplete filling 1 (*Gif1*). The rice Dense and erect protein 1 (*dep1*) is the major rice grain yield QTL that determines the panicle architecture. The locus was first identified by two independent research groups with quantitative trait locus analysis to control grain yield, grain number per panicle and panicle morphology (Huang et al., 2009; Wang et al; 2009; Zhou et al., 2009). *Dep1* encodes plant-specific gamma (γ) subunit of GTP-binding protein (Chakravorty et al., 2011). Grain number is one of the most important traits in the determination of crop productivity. A pioneering study by Ashikari et al. (2005) on the molecular analysis of grain number determination in rice identified grain number 1a (*Gn1a*) as a major QTL of grain number using the backcross inbred lines (BILs) of Koshihikari (lower grain number) and Habataki (higher number). *Gn1a* encodes an enzyme, namely cytokinin oxidase/dehydrogenase (Osckx2), which degrades the phytohormone Cytokinin (Galuszka et al., 2001). Grain size 3 (*Gs3*) was the first major QTL in this category to be isolated, and it has been found that it contributes to both grain length and weight. *Gs3* was identified in progeny derived from a cross between Minghui 63 (large grain) and Chuan 7 (small grain) alleles (Fan et al., 2006). Grain weight and width (*Gw2*) influences grain width and weight. This QTL was identified in progeny from a cross between WY 3 (large grain) and Fengaizhan-1 (small grain). *Gw2* encodes the ring-type protein with E3 ubiquitin ligase activity that localizes to the cytoplasm and is constitutively expressed in various tissues, which is known to be involved in the degradation processes of the ubiquitin-proteasome pathway (Song et al., 2007). Grain incomplete filling 1 (*Gif1*) encodes a cell wall invertase required for carbon partitioning during early grain filling (Wang et al., 2008).

This study was aimed at determining the presence of five major yield responsive genes (*Gif1*, *Gn1a*, *Gw2, Gs3* and *Dep1*) shown to be among key determinants of rice yield in ten rice cultivars (Faro 44, Faro 45, Faro 60, Faro 61, Dan-boto, Dan-kaushi, Jan-iri, Dan rai-rai, Bakin-iri and Walkin-kambari) grown in Nigeria using candidate gene SSRS markers (Cg-SSRs).

MATERIALS AND METHODS

Samples

A total of ten rice genotypes including four improved rice (high yielding) genotypes obtained from National Cereal Research

Table 1. SSRs found in *Dep1*, *Gn1a*, *Gs3*, *Gw2* and *Gif1* with minimum of five (5) di- and tri- motif repeats.

S/N	Gene	Motif	Number of repeats	SSR start	SSR end
		CT	5	9	18
1	*Dep1*	GC	5	181	190
		TC	5	2204	2213
		CA	5	105	114
		CA	5	3613	3622
2	*Gn1a*	CGC	5	375	389
		TCC	5	1205	1219
		TAA	5	4070	4084
		AT	5	4243	4252
3	*Gs3*	TCC	6	5797	5814
		CTC	6	5865	5882
		GCA	9	155	181
4	*Gw2*	GGA	5	494	508
		CGC	5	735	749
5	*Gif1*	TAG	6	956	973

The primer pair (forward and reverse) were designed on the sequences flanking at least one of the following motif repeats for each gene (SSRs gene markers). *Dep1 = dense and erect protein 1*; *Gif1 = grain incomplete filling 1*; *Gn1a = grain number 1a*; *Gs3 = grain size 3*; *Gw2 = grain weight* and *width*; SSR = simple sequence repeats.

Institute (NCRI), Badeggi and six local rice (low yielding) genotypes obtained from rice farmers within Argungu and Suru towns of Kebbi State were collected. The improved rice genotype samples included: Faro 44, Faro 45, Faro 60, and Faro 61. The local rice genotype samples included: Dan-boto, Dan-kaushi, Jan-iri, Dan rai-rai, Bakin-iri and Walkin-kambari. The seeds were stored in 9' ×4' brown envelops that allowed for aeration and stored in refrigerator (4°C) before used.

Samples planting

The seeds were soaked for 2 h prior to planting, then an average of ten seeds was each transferred into individual sterilized petri-dishes bedded with filter-papers. The seeds were watered everyday by way of soaking the filter-papers with few drops of water. This was maintained for seven days after which they were harvested and stored in the refrigerator (-20°C) before use for DNA extraction.

Genomic DNA extraction using cetyltrimethylammonium bromide (CTAB)

The extraction protocol used in this study was adopted from Turaki et al. (2017), but with modification that involves the removal of phenol:chloroform: isoamylalcohol (25:24:1) step and replace the step with chloroform: isoamylalcohol (24:1). The removal of phenol is associated with its toxicity and difficulty associated with disposing of its waste.

Estimation of quantity and quality of isolated genomic DNA using UV/Visible Nano-drop Spectrophotometer

The re-suspended total nucleic acid was allowed to stand for at least 12 h in refrigerator (-20°C) to allow it dissolve evenly in the solution, then diluted in molecular grade water in a ratio of 1:10 before taking the absorbance readings. The quantity and purity of the extracted total nucleic acid were analysed using Nano-drop Spectrophotometer, as total nucleic acid purity is the ratio of spectrophotometric absorbance of total nucleic acid at $\lambda=260$ nm and protein at $\lambda=280$ nm. First, 2 µl of molecular graded water was used to set a reference value (blank) then absorbance of the samples was measured. 260/280 absorbance ratio of the samples was recorded.

SSRs mining

The gene fasta format reference sequences retrieved from NCBI database were copied and pasted into SSR identification tool (SSR IT) according to Temnykh et al. (2001). The criteria for the search were that only SSRs with di- and tri-nucleotide motifs, having minimum of five number of repeats were mined (Table 1).

Primers design

Both forward and reverse primers were designed manually, one set for each gene. The primers were designed to sit on the flanking sequences of the identified SSRs region. Vector NTI (version 11.5 advanced) (http://thermofisher.com/ng/en/home/life-science/cloning/vector-nti-advance-download.html) software and oligocalc northwestern (http://basicbiotools.nubic.northwestern.edu/oligocalc.html) software were used to test for the following parameters: primer length 18 to 25 base pairs (bp), melting temperature (T_m), 50 to 60°C, forward and reverse primers T_m difference of not more than 4°C, percentage GC content 45 to 65 and expected amplicon size of 150 to 500 bp (Table 2).

Table 2. Yield responsive gene, respective primers with %GC content and Tm, expected band sizes, motif(s) with repeat number and location of motifs within gene sequence.

Gene	Forward primer, 5'→3' (%GC/Tm)	Reverse primer, 5'→3' (%GC, Tm)	Expected (Amplicon size, bp)	Motif repeats (captured)	Location of motifs (Within gene)
Dep1	CTCCTCATCGCATCGCAT (55.6, 51.5)	GCCAACTCAGAACCACCC (61.1, 49.7)	236	$(GC)_5$	5'UTR
Gn1a	GCCTTCCATCGTCAGCAC (61.1, 50.9)	GCAGTTGAGCATGAGGAGC (57.9, 50.5)	185	$(CA)_5$	5'UTR
Gw2	CCTCCATCTCCACCACCGTA (60.0, 54.5)	GGAGGAAGTGAGGAAGAGGG (60.0, 51.7)	408	$(GGA)_5$ and $(CGC)_5$	5'UTR and CDS
Gs3	CAAGTGCGTGCTGCCTCA (61.1, 53.9)	AGCGGCACGAGCATCAC (64.7, 52.5)	390	$(TCC)_6$ and $(CTC)_6$	CDS (both)
Gif1	GCTGCTGCGTCATCAACATT (50.0, 53.6)	GTGGTCATCAGCCGTGGTAG (60.0, 52.7)	285	$(TAG)_6$	Intron

Bp, Base pair; CDS, coding sequences; *Dep1, dense and erect protein 1*; *Gif1, grain incomplete filling 1*; *Gn1a, grain number 1a*; *Gs3, grain size 3*; *Gw2, grain weight and width*; Tm melting temperature; UTR, untranscribed region.

Constitution of primers

A set of five primers (forward and reverse) obtained from Sigma-aldrich® (sigma-life sciences, UK) were used for the PCR analysis. The tubes containing the primers were first centrifuged to pull down primers that might have been displaced from the bottom of the tubes during shipment. Primer stock solutions were prepared according to manufacturer's instruction with nuclease free distilled water. The stock solution was divided into smaller 25 µl aliquots for long term storage to avoid frequent freeze-thaw cycles and accidental cross contamination.

Polymerase chain reaction (PCR)

The PCR reaction was set using a total of 40 µl reaction mixture containing 4 µl 10x dream Taq green buffer (20 mm $MgCl_2$, Thermo Scientific, UK), 1.0 µl of dNTPs (2 mm of each dNTP, Thermo scientific, UK), 2.0 µl of each primer (0.1 µm of each primer, Sigma-life Sciences, UK), 0.4 µl of dream Tag total nucleic acid polymerase (1.25 U, Thermo Scientific, UK), 2 µl total nucleic acid template and 28.6 µl of molecular grade water. Thermal-cycling was performed in a supercycler (SC300, kyratec, Queensland, Australia) programmed as follows: 95°C/3 min initial denaturation, 30 cycles of 94°C/20 s, 58°C/30 s (for *Gw2*, *Gif1* and *Gs3*), 52°C/30 s (for *Dep1* and *Gn1a*) and 72°C/1 min and final cycle of 72°C/5 min.

Agarose gel electrophoresis

Amplification products were resolved on 1% agarose gel, stained with 5 µl ethidium-bromide in a horizontal electrophoresis tank system (Flowgen biosciences, UK) containing 1x TBE buffer. About 6 µl of 100 bp ladder (promega, UK) were used at both edges. The electrophoresis run was programmed as follows: time/75 min, current/80 milliamps (mA) and voltage/120 V. The gels were visualized and their snap shots taken in a gel documentation system (desktop gel imager, scope 21).

Sequence analysis

A total of four PCR-amplicons were sequenced by Inqaba biotech, South Africa (www.inqababiotec.co.za) using Sanger's sequencing method (Sanger et al., 1977). Two amplicons each from the two gels that gave strong and targeted bands (Figure 1a and c) were chosen, one amplicon each from a local cultivar and an improved cultivar. The obtained sequences (chromatograms) were viewed using FinchTV (www.digitalworldbiology.com/finchtv). Molecular evolutionary genetic analysis (mega7.0) software (www.megasoftware.net/mega.php) was used to edit the obtained sequences. The edited fasta format sequences were used for similarity searches using the basic local alignment search tool (BLAST) program of Molecular Evolutionary Genetic Analysis (MEGA7.0) (Kumar et al., 2016) software in the National Centre for Biotechnology Information (NCBI) genbank databases (www.ncbi.nlm.nih.gov/blast.cgi). The evolutionary history was inferred using the neighbor-joining method (Saitou and Nei, 1987). The bootstrap consensus tree inferred from 1,000 replicates is taken to represent the evolutionary history of the taxa analyzed (Felsenstein, 1985). The evolutionary distances were computed using the maximum composite likelihood method (Tamura et al., 2004). All positions containing gaps and missing data were eliminated.

RESULTS

Spectrophotometric absorbance

The ratio of spectrophotometric absorbances of the extracted total nucleic acid at λ=260 nm and protein at λ=280 nm were between 1.90 and 2.20, with a mean value of 2.05. The quantity of the extracted ranged between 125.2 and 253.5 ng/µl, with a mean value of 179.8 ng/µl.
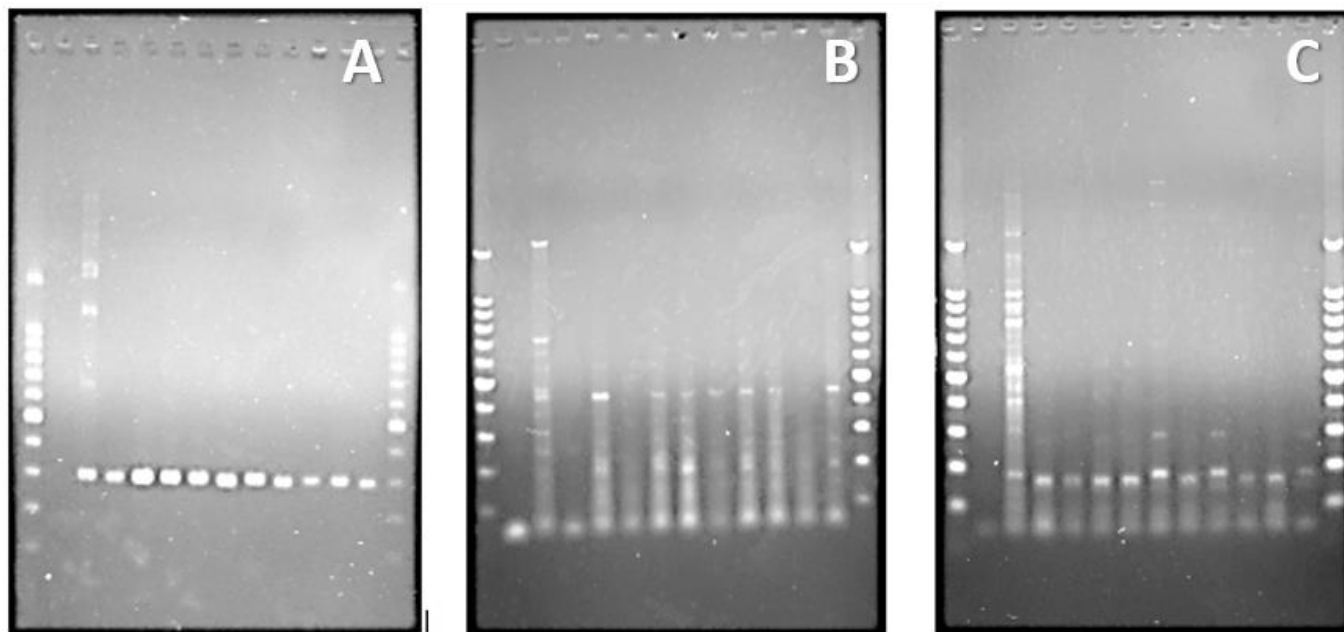
Figure 1. (A) Electrophoresis of rice *Gif1* gene. Lane M = 100 bp total nucleic acid ladder (Promega; USA); lane 1 = negative control (no template control); lane 2 = positive control (Nipponbare cultivar); lane 3 = Faro 44; lane 4 = Faro 45; lane 5 = Faro 60; lane 6 = Faro 61; lane 7 = Dan-boto; lane 8 = Bakin-iri; lane 9 = Jan-iri; lane 10 = Dan-rairai; lane 11 = Walkin-kambari; lane 12 = Dan-kaushi. **(B)** Electrophoresis of rice *Gw2* gene. Lane M = 100 bp total nucleic acid ladder (promega; USA); lane 1 = negative control (no template control); lane 2 = positive control (Nipponbare cultivar); lane 3 = Faro 44; lane 4 = Faro 45; lane 5 = Faro 60; lane 6 = Faro 61; lane 7 = Dan-boto; lane 8 = Bakin-iri; lane 9 = Jan-iri; lane 10 = Dan-rairai; lane 11 = Walkin-kambari; lane 12 = Dan-kaushi. **(C)** Electrophoresis of rice *Gn1a* gene. Lane M = 100 bp total nucleic acid ladder (Promega; USA); lane 1 = negative control (no template control); lane 2 = positive control (Nipponbare cultivar); lane 3 = Faro 44; lane 4 = Faro 45; lane 5 = Faro 60; lane 6 = Faro 61; lane 7 = Dan-boto; lane 8 = Bakin-iri; lane 9 = Jan-iri; lane 10 = Dan-rairai; lane 11 = Walkin-kambari; lane 12 = Dan-kaushi.

SSRs mining

The SSRs mined using SSR identification tool (SSR IT) satisfying the di- and tri-nucleotide motifs with minimum of five number of repeats are shown in Table 1.

Primers designing

Five set of primers, one each for a gene on the flanking sequences of at least one SSR were designed. The details of the genes, their respective primers, percentage GC content (%GC), melting temperature (T_m), expected band sizes, motif(s) with repeat number and location of motifs within gene sequence are shown in Table 2.

PCR analysis

In this study, 1% agarose gels were used to analyze the amplified PCR products and the expected band sizes were in the range of 185 to 408 bp. The amplified total nucleic acid product were stained with a chemical dye (0.1% ethidium bromide) which intercalates between the two strands of the duplex facilitating a large increase (up to 1,000 times) in their capacity to fluoresce under UV light for visualization. Three out of the five gels showed PCR products of expected band sizes. The remaining two showed no bands on the sample lanes while the positive controls showed multiple bands on both gels but not of the expected sizes:

(1) *Gif1* gene: The positive control (lane 2) and all the samples (lanes 3 to 12) showed single, clear and expected band sizes of approximately 285 bp. The negative control (lane 1) showed no band as expected (Figure 1a).
(2) *Gw2* gene: It can be seen from Figure 1b that the positive control lane (2) together with samples lanes 4, 6, 7, 8, 9, 10, 11 and 12 all showed expected band sizes of approximately 400 bp on the gel. But lanes 3 and 5 show no band. The negative control (lane 1) showed no band as expected (Figure 1b).
(3) *Gn1a* gene: The positive control (lane 2) and all the samples (lanes 3 to 12) showed single, clear and expected band sizes of approximately 185 bp. The negative control (lane 1) showed no band as expected (Figure 1c).
(4) *Gs3* gene: All the sample lanes showed no band (lanes 3 to 12). The positive control showed multiple

bands and not of the expected band sizes of approximately 390 bp. The negative control (lane 1) showed no band as expected (not shown).

(5) *Dep1* gene: All the sample lanes (lane 3 to 12) showed no band. The positive control (lane 2) showed multiple bands and not of the expected band sizes of approximately 236 bp. The negative control showed no band as expected (not shown).

Sequence analysis

The edited fasta format of *Gif1* (Faro 44 and dan boto cultivars) and *Gn1a* (Faro 60 and Dan rairai cultivars) partial gene sequences are as follows:

(1) Faro 44 - *Gif1* partial gene sequence:

TTTATAATATATCCTGTCACGAATCTCTGTCTACTAGT
AGTAGTAATAGTACTAGAACTTTTATGCCTTGCAACTT
GCAATTTCGTTGTACGGGAGAGAACTGTAGTTAGTG
ACGCCTTTCATGGTACGATTAAAGGTTCAAAGCACAT
TTTAGCACGAAAATGGTAGGCGCACTGGGACTCCAC
ATGCAGGCTTGCTTGTCGACCGTGGGGTACCTAGCC
CCTACCACGGCTGATGACCAC

(2) Dan boto - *Gif1* partial gene sequence:

TTTATAATATATCCTGTCACGAATCTCTGTCTACTAGT
AGTAGTACTAGAACTTTTATGCCTTGCAACTTGCAAT
TTCGTTGTACGGGAGAGGACTGTAGTTAGTGACGCC
TTTCATGGTAGGATTAAAGGTTCAAAGCACATTTTAG
CACGAAAATGGTAGGCGCACTGGGACTCCACATGCA
GGCTTGCTTGTCGACCGTGGGGTACCTAGCCCCTAC
CACGGCTGATGACCACAAC

(3) Faro 60 - *Gn1a* partial gene sequence:

CACCTTGTCCCTTCTACAATGGTGCAAGAACACACAA
ATTCACACACACACTGACACACACAAACCGATCGATT
GATTGATTGATAATGAAGCAAGAGCAGGTCAGGATG
GAGTGCTCCTCATGCTCAACTGC

(4) Dan rairai - *Gn1a* partial gene sequence:

TGTCCCTTCTAAATGGTGCTCGAACACACAAATTCAC
ACACACACTGACCACACAAACCGATCGATTGATTGAT
TGATAATGAAGCAAGAGCAGGTCAGGATGGCAGTGC
TCCTCATGCTCAACTGC

BLAST search

For *Gif1* gene, Faro 44 (an improved cultivar) and Dan-boto (a local cultivar) were sequenced. The BLAST search for *Gif1* sequence from Faro 44 shows 100% sequence identity with *Oryza sativa* indica (cultivar: RP

bio 226) and Bac clone: Osigb0134p10 sequences. The sequence also shows 98% sequence identity with *O. sativa* japonica (cultivar: Nipponbare), *O. sativa* indica (cultivar: Shuhui 498) and *O. sativa* Bac clone: Oj000126_13. Comparing the obtained sequence from Dan-boto shows 97% sequence identity with *O. sativa* indica (cultivar: RP bio 226) and Bac clone: Osigb0134p10 sequences. The sequence also shows 96% identity with *O. sativa* japonica (cultivar: Nipponbare), *O. sativa* indica (cultivar: Shuhui 498) and *O. sativa* Bac clone: Oj000126_13. For *Gn1a*, Faro 60 (an improved cultivar) and Dan-rai-rai (a local cultivar) were sequenced. Comparing the obtained sequence from Faro 60 with *O. sativa* japonica cytokinin dehydrogenase 2 mRNA (loc 4327333), *O. sativa* japonica (cultivar: Nipponbare), *O. sativa* japonica Bac clone: B1046g12 and *O. sativa* japonica Pac clone: P0419b01 shows 99% sequence identity. The sequence, however, shows 87% sequence identity with *O. sativa* indica Shuhui 498 and *O. sativa* indica RP bio 226. Comparing the obtained sequence from Dan-rai-rai with *O. sativa* japonica cytokinin dehydrogenase 2 mRNA (loc 4327333), *O. sativa* japonica (cultivar: Nipponbare), *O. sativa* japonica Bac clone: B1046g12 and *O. sativa* japonica Pac clone: P0419b01 shows 96% sequence identity. An 86% sequence identity was obtained upon comparing the sequence obtained from Dan-rai-rai with that of *O. sativa* indica Shuhui498 and *O. sativa* indica RP bio 226.

Phylogenetic analysis

Phylogenetic analysis of Gif1 gene

Phylogenetic analysis of the seven partial nucleotide sequences of rice *Gif1* gene showed two major clusters (A and B). Bac clone: Oj000126_13 cluster separately (cluster B) while the remaining six cultivars (Faro 44, Dan boto, RP bio-226, Nipponbare, Shuhui 498, Bac clone: Osigb0134p10) cluster together (cluster A). Cluster A was sub-grouped into two clusters. Five of the six cultivars cluster together (cluster I) while Nipponbare cultivar clusters differently. Cluster I was equally divided into two clusters. Shuhui 498 forms an independent cluster while the remaining four cultivars cluster together (cluster II). In the subsequent grouping, Dan boto (a local cultivar) forms a cluster of its own while the remaining three forms cluster III. Lastly, RP bio 226 and Bac clone: Osigb0134p10 cluster together (cluster IV) while Faro 44 (an improved cultivar) forms a different cluster (Figure 2A).

Phylogenetic analysis of Gn1a gene

Phylogenetic analysis of the seven partial nucleotide

sequences of rice *Gn1a* gene shows two major clusters (cluster A and B). Cluster A has RP bio-226, Nipponbare, Shuhui 498, Faro 60, Dan rairai and Pac clone: P0419b01 while Bac clone: b1046g12 cluster differently (cluster B). Cluster A was grouped into two clusters (I and II), each cluster having three cultivars. Cluster I was re-grouped into two clusters; cluster I (RP bio-226 and Shuhui 498) while Faro 60 forms a different cluster. Cluster II was equally re-grouped into two clusters; cluster II (Pac clone: P0419b01 and Dan rairai) while Bac clone: b1046g12 forms a different cluster (Figure 2b).

DISCUSSION

Extraction of pure total nucleic acid free from contaminants is the first critical step for molecular biology experiments. Contaminants such as polyphenols and polysaccharides are often co-precipitated with total nucleic acid (Mumford and Seal, 1997). In the present study, the extraction method was based on modified CTAB method reported by Turaki et al. (2017). Young fresh leaf tissues were used for the total nucleic acid extraction to avoid the potentials of nucleic acid contamination by plant metabolites that interfere with the solubilization of precipitated nucleic acids (Puchooa, 2004). Another advantage of using young leaf is the ease of the disruption of cell wall and membranes to release nuclear materials.

The simplest and fastest method of determining total nucleic acid concentration and purity is the spectrophotometric method (Teare et al., 1997). The best test for total nucleic acid quality is its ability to be amplified in downstream analysis. It is a common knowledge that total nucleic acid absorbs maximally at 260 nm while proteins absorb at 280 nm wavelength. Total nucleic acid purity is determined by its absorbance ratio of 260/280 nm wavelength (Glasel, 1995). By calculating absorbance ratio of 260/280 nm wavelengths, it can be determined whether total nucleic acid preparations are contaminated with protein or not. Extracted total nucleic acid with protein contamination absorb strongly at 280 nm due to aromatic rings on amino acid side chains and will lower the ratio to <1.8. If the ratio is greater than or equal to 1.8 ($\geq$ 1.8) the total nucleic acid can be considered as pure total nucleic acid (William et al., 1997). Pure RNA has a ratio of approximately 2.0.

The ratios in the present study fall within a range of 1.90 and 2.20. Thus, the total extracted total nucleic acid can be considered good for downstream analysis (that is, 2.0±0.20). The total nucleic acid concentrations of the diluted samples were of reasonable quantity (yield range between 15.50 and 25.35 ng/µl).

In this study, 1% agarose gels were used to analyze the amplified PCR products and the expected band sizes were in the range of 185 to 408 bp. The amplified total

nucleic acid products were stained with a chemical dye (0.1% ethidium bromide) which intercalates between the two strands of the duplex facilitating a large increase (up to 1,000 times) in their capacity to fluoresce under UV light for visualization (Srinivasan et al., 1993). The band size of ~285 bp PCR products for *Gif1* and ~185 bp *Gn1a* genes on the gels (Figure 1a and c, respectively) shows single and clear bands for all the rice genotypes screened. This indicates that all the genotypes have both *Gif1* and *Gn1a* genes. This is supported by the expected band sizes of 285 and 185 bp for *Gif1* and *Gn1a* PCR positive results for the positive control genotype. The PCR assay did work because the no total nucleic acid-template control (negative control) lanes showed no band as expected.

In the present study, *Gw2* gene was also studied. Eight genotypes (Faro 45, Faro 61, Dan-boto, Bakin-iri, Jan-iri, Dan-rai-rai, Walkin-kambari and Dan-kaushi) show weak bands of approximately 400 bp. The weak bands may be due to small copy numbers of the targeted DNA region within the extracted genomic total nucleic acid. Lorenz (2012) pointed out that it is not the concentration of the extracted total nucleic acid that actually counts rather the number of copies of the region targeted. He posited that copies of 10^4 to 10^7 of the targeted total nucleic acid region are required for PCR with total reaction volume of 50 µl. The two other sample genotypes (Faro 44 and Faro 60) show no bands at all. This indicates that the two genotypes have no *Gw2* gene or have a different variant of the gene resulting in primer mis-match with the complementary sequences. The puzzles can be solved by using degenerate primer sets which have the ability to amplify different variants of a gene instead of the sequence specific primers used in this research (Gahoi et al., 2013). Furthermore, primer sets used to screen for *Gs3* and *Dep1* (not shown) require further optimization because none of the genotypes screened was amplified and the positive controls showed multiple unexpected bands. These indicate that both the primer sets may be having regions of preference on the genomic total nucleic acid different from the targeted sequence on the positive control genotype. Moreover, success of PCR is critically dependent on the design of an effective primer pair (Kalendar et al., 2009). PCR failures could be as result of polymorphisms such as SNP, indels and copy number variations (Piriyaponga et al., 2009). Unexpected SNP in a designed primer, in particular in the 3' end (SNP-in primer), primers designed within the intron/exon boundaries or within repetitive DNA elements are possible reasons for PCR failures (Piriyaponga et al., 2009).

The best standard in PCR product analysis is direct sequencing of the amplicon to determine its actual nucleotide composition. For the Faro 44 sequence (obtained in the present study), 100% sequence identity was obtained with *O. sativa* indica (cultivar: RP bio 226) and Bac clone: Osigb0134p10 sequences. This shows no nucleotide deletion/insertion or mismatching between

the sequences. The difference (98%) observed in the sequence identity with *O. sativa* japonica (cultivar: Nipponbare), *O. sativa* indica (cultivar: Shuhui 498) and *O. sativa* Bac clone: Oj000126_13 is as a result of triplet base deletion of 'TAG' in the Faro 44 sequence at the positions 956, 957 and 958. The 'TAG' deleted is the first of the six 'TAG' motif SSR repeats in the gene. For the Dan-boto sequence, the difference in the sequence identity (97%) with *O. sativa* indica (cultivar: RP bio 226) and Bac clone: Osigb0134p10 sequences is as a result of two triplet base deletion of 'TAG' in the Dan-boto sequence at positions 956, 957, 958, 960, 961 and 962. The 'TAGs' deleted are the first and second of the six 'TAG' motif SSR repeats in the gene. The sequence also shows 96% identity with *O. sativa* japonica (cultivar: Nipponbare), *O. sativa* indica (cultivar: Shuhui 498) and *O. sativa* Bac clone: Oj000126_13. The difference was also consequent to deletion of the first three 'TAGs' of the six 'TAG' motif SSR repeats. For *Gn1a* gene, the 99% sequence identity of Faro 60 with *O. sativa* japonica cytokinin dehydrogenase 2 mRNA (Loc4327333), *O. sativa* japonica (cultivar: Nipponbare), *O. sativa* japonica Bac clone: B1046g12 and *O. sativa* japonica Pac clone: P0419b01 is as a result of deletion of 'C' nucleotide at 176 position in the Faro 60 obtained sequence. The sequence, however, shows 87% sequence identity with *O. sativa* indica Shuhui498 and *O. Sativa* indica RP bio 226. The lower identity score is as a result of 16 nucleotide deletion ("CCGATCGATTGATTGA" at 129 to 144 positions) in the sequences of the two accessions (*O. sativa* indica Shuhui498 and *O. sativa* indica RP bio 226) which are conserved in the Faro 60 sequence. The obtained sequence from Dan-rai-rai shows 96% sequence identity with *O. sativa* japonica cytokinin dehydrogenase 2 mRNA (Loc4327333), *O. sativa* japonica (cultivar: Nipponbare), *O. sativa* japonica Bac clone: B1046g12 and *O. sativa* japonica Pac clone: P0419b01. The percentage identity obtained is as a result of insertion of an 'A' nucleotide at position 64, deletion of 'C' at position 82 and another deletion of 'A' nucleotide at position 120 within the sequence obtained from Dan-rai-rai. An 86% sequence identity was obtained upon comparing the sequence obtained from Dan-rai-rai with that of *O. sativa* indica Shuhui498 and *O. sativa* indica RP bio 226. This lower identity score is as a result of insertion of an 'A' nucleotide at position 64, deletion of 'C' at position 82, deletion of 'A' nucleotide at position 120 and another 16 nucleotide deletion ("CCGATCGATTGATTGA" at 129 to 144 positions) in the sequences of the two accessions (*O. sativa* indica Shuhui498 and *O. sativa* indica RP bio 226).

The two dendrograms (Figure 2a and b) demonstrate clearly all the eleven cultivars (positive control cultivar inclusive) shared ancestral relationships. The dendrograms also show both genetic similarities and differences abound among the rice cultivars. However, a dendrogram from complete gene sequences would have been more reliable because it will take into account the differences in the entire gene sequences. Notwithstanding, the similarities observed in the dendrograms indicate genetic, structural and functional relationships. The differences indicate the genetic diversity of both *Gif1* and *Gn1a* genes among the rice cultivars. The genetic diversity indicates possibilities for genetic improvement of rice yield through appropriate selection and cross breeding.

Conclusion

In this research work, five sequence specific candidate gene SSRs based (CgSSRs-based) primers have been designed. Genomic total nucleic acids of good quality and concentration were extracted using optimized CTAB method. PCR products revealed the presence of *Gif1* and *Gn1a* genes in the ten rice cultivars screened. Eight rice cultivars (Faro 45, Faro 61, Dan boto, Dan kaushi, Dan rai-rai, Jan iri, Bakin-iri and Walkin-kambari) have *Gw2* gene while the remaining two cultivars (Faro 44 and Faro 60) do not have the gene. The primers for *Gs3* and *Dep1* need further PCR optimization or new degenerate primers be designed. It can be deducted from the two dendrogram that both genetic similarities and differences are abound in both *Gif1* and *Gn1a* genes. Based on the aforementioned findings, it can be concluded that candidate gene SSRs-based (CgSSRs-based) markers are valuable molecular markers for the identification of major yield responsive genes in rice.

RECOMMENDATIONS

Based on the findings of this study, the following recommendations are made:

(1) Research should be conducted to ascertain the presence or otherwise of other yield responsive genes not covered by this research (such as *Gs5*, *Gs6*, *Gs2*, *Gw5*, *Ipa1*, *DST*, *Glw7*) using candidate gene SSRs (CgSSRs) and other molecular markers.
(2) Those yield responsive genes found to be positive should be cloned and sequenced in order to know the level of variation throughout the genes.
(3) Quantitative RT-PCR should be carried out to know the level of expression of the genes in those cultivars found to be positive of the genes screened and those found to be robustly expressed be transferred to other accessions for their yield improvement.

CONFLICT OF INTERESTS

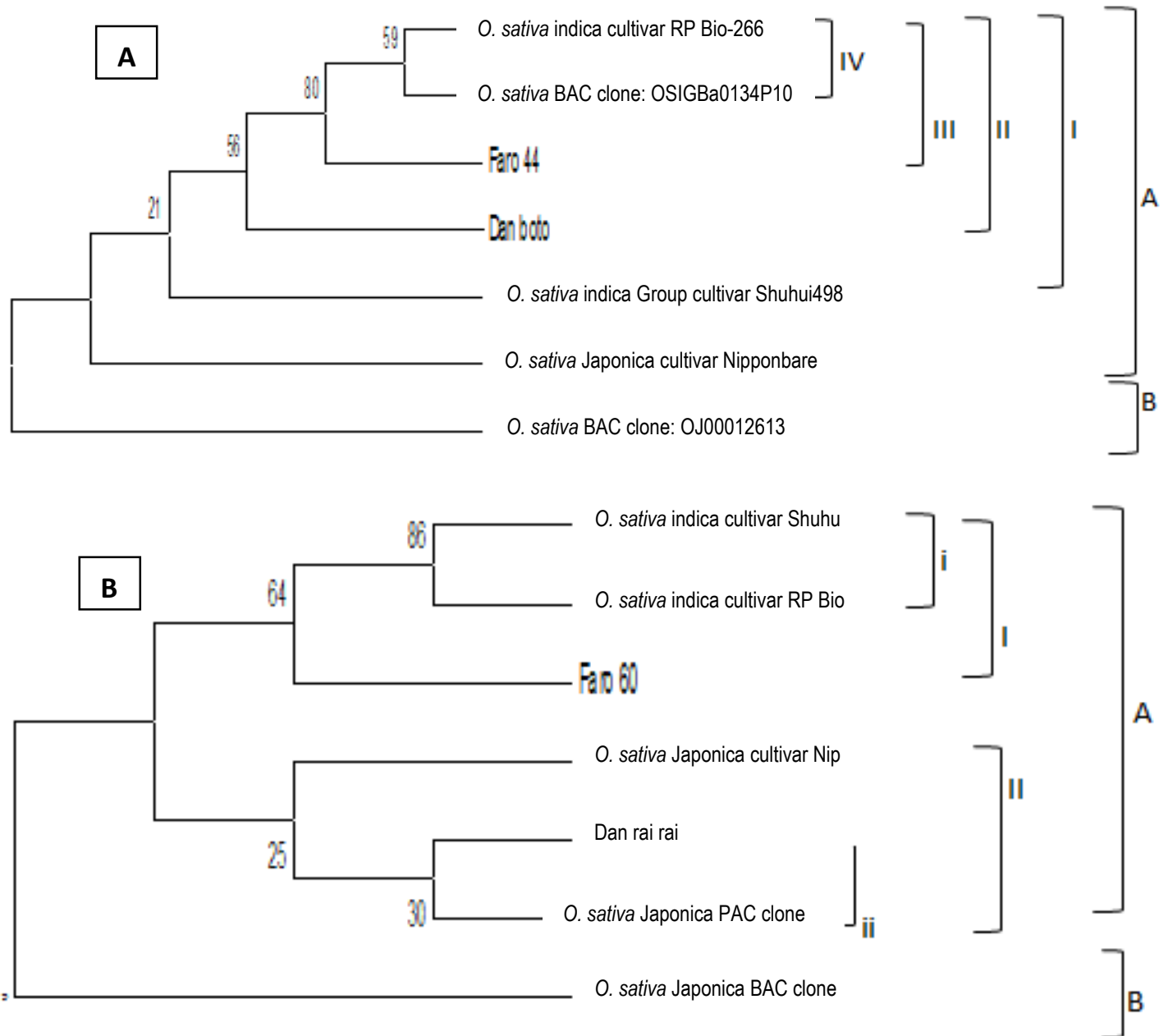The authors have not declared any conflict of interests.

Figure 2. (A) Bootstrap consensus neighbor-joining phylogenetic tree (dendrogram) generated from nucleotide sequences of 7 rice *Gif1* partial gene sequences. **(B)** Bootstrap consensus neighbor-joining phylogenetic tree (dendrogram) generated from nucleotide sequences of 7 partial *Gn1a* gene sequences.

REFERENCES

Ashikari M, Sakakibari H, Lin S, Matsuaka M (2005). Cytokinin Oxidase regulates rice grain production. Science 309(5735):741-745.
Chakravorty D, Trusov Y, Zhang W, Acharya BR, Sheahan MB, McCurdy DW, Assmann SM, Botella JR (2011). An atypical

heterotrimeric G-protein gamma-subunit is involved in guard cell K$^{(+)}$-channel regulation and morphological development in *Arabidopsis thaniala*. Plant Journal 67:840-851.

Chan KF, Sun M (1997). Genetic diversity and relationships detected by isozyme and rapd analysis of crop and wild species of *Amaranthus*. Theoretical and applied genetics 95:865-873.

Cheung F (2014). Rice. Nature outlook 514:S61.

Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, Li X, Zhang Q (2006). *Gs3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theoretical and applied genetics 112:1164-1171.

Felsenstein J (1985). Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783-791.

Gahoi S, Arya L, Anil R, Marla Ss (2013). Dpprimer – a degenerate PCR primer design tool. Bioinformation 9(18):937-940.

Galuszka P, Frebort I, Sebela M, Sauer P, Jacobsen S, Pec P (2001). Cytokinin oxidase or dehydrogenese? Mechanism of cytokinin degradation in plants. European Journal of Biochemistry 268:450-461.

Glasel JA (1995). Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. Biotechniques 18(1):62-63.

Govindaraj M, Vetriventhan M, Srinivasan M (2015). Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. Genetics Research International 1:1-15.

Hu J, Wang Y, Fang Y, Zeng L, Xu J, Yu H (2015). A rare allele of *Gs2* enhances grain size and grain yield in rice. Molecular plant 8:1455-1465.

Huang X, Qian Q, Liu Z, Sun H, He S, Luo D, Xia G, Chu C, Li J, Fu X (2009). Natural variation at the *Dep1* locus enhances grain yield in rice. Nature genetics 41(4):494-497.

Kalendar R, Lee D, Schulman AH (2009). Fastpcr software for pcr primer and probe design and repeat search. Genes, genomes and genomics 3:1-14.

Kumar S, Stecher G, Tamura K (2016). Mega7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular biology and evolution 33:1870-1874.

Li S, Zhao B, Yaun D, Qian Q, Tang L, Wang B, Liu X, Zhang J, Wang J, Sun J, Liu Z, Feng YQ, Yuan L, Li C (2013). Rice zinc finger protein DST enhances grain production through controlling Gn1a/Osckx2 expression. Proceedings of the national academy of science 110(8):3167-3172.

Li Y, Fan C, Xing Y, Jiang Y, Luo L, Sun L (2011). Natural Variation In *Gs5* Plays An Important Role In Regulating Grain Size And Yield In Rice. Nature genetics 43:1266-1269.

Lorenz TC (2012). Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. Journal of visualized experiments 63:3998.

Miura K, Ikeda M, Matsubara A, Song XJ, Ito M, Asano K, Matsuoka M, Kitano H, Ashikari M (2010). *Osspl14* promotes branching and higher grain productivity in rice. Nature genetics 42:545-549.

Mumford RA, Seal SE (1997). Rapid single-tube immunocapture RT-PCR for the detection of two yam potyviruses. Journal of virological methods 69:73-79.

Piriyaponga J, Ngamphiw C, Assawamakin A, Wangkumhang P, Suwannasri P, Ruangrit U, Agavatpanitch G, Tongsima S (2009). Rexprimer: an integrated primer designing tool increases PCR effectiveness by avoiding 3' SNP-in-primer and mis-priming from structural variation. BMC genomics 10(3):S4.

Premium times newspaper (2015). Nigeria spends $ 2.41 billion on rice importation in three years. Available: www.premiumtimesng.com/business/187406-nigeria-spends-2-4bn-on-rice-importation-in-3-years-emefiele.html via the internet. Accessed 2016, Jul 13.

Puchooa D (2004). A simple, rapid and efficient method for the extraction of genomic DNA from lychee (*litchi chinensis*sonn.). African Journal of Biotechnology 3(4):253-255.

Punch newspaper (2017). Rice production in Nigeria increases to 5.8 million tonnes in 2017. Available: www.punch.com/rice-production-in-nigeria-increases-to-5.8-m-tonnes-in-2017-rifan/ampvia the internet. Accessed 2017, Nov. 18.

Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution 4(4):406-425.

Sanger F, Nicklen S, Coulson AR (1977). DNA sequencing with chain-terminating inhibitors. Production of the National Academy of Science, USA. 74:5463-5467.

Si L, Chen J, Huang X, Gong H, Luo J, Hou Q (2016). Osspl13 controls grain size in cultivated rice. Nature genetics 48:447-456.

Song XJ, Huang W, Shi M, Zhumz, Lin HX (2007). A QTL for rice grain width and weight encodes a previously unknown ring-type E3 ubiquitin ligase. Nature genetics 39:623-630.

Srinivasan K, Morris SC, Girard JE, Kline MC, Reeder DJ (1993). Enhanced detection of PCR products through use of TOTO and YOYO intercalating dyes with laser-induced fluorescence–capillary electrophoresis. Applied and theoretical electrophoresis 3(5):235-239.

Sun L, Li X, Fu Y, Zhu Z, Tan L, Liu F, Sun X, Sun X, Sun C (2013). Gs6, a member of the gras gene family, negatively regulates grain size in rice. Journal of integrative plant biology 55:938-949.

Tamura K, Nei M, Kumar S (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the national academy of sciences, USA 101:11030-11035.

Temnykh S, Declerck G, Lukashova A, Lipovich L, Cartinhour S, Mccouch S (2001). Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome resources 11(8):1441-1452.

Teare JM, Islam R, Flanagan R, Gallagher S, Davies MG, Grabou C (1997). Measurement of nucleic acid concentration using the dynaquanttm and genequanttm. Biotechniques 22:1170-1174.

Turaki AA, Ahmad B, Magaji UF, Abdulrazak UK, Yusuf BA, Hamza AB (2017). Optimized cetyltrimethylammonium bromide (ctab) DNA extraction method of plant leaf with high polysaccharide and polyphenolic compounds for downstream reliable molecular analyses. African Journal of Biotechnology 16(24):1354-1365.

Vanguard Newspaper (2013). Nigeria: rice consumption and importation in Nigeria, no solution in sight. All Africa archieve. Available: www.allafrica.com/stories/201312300524.html via the internet. Accessed 2016 Aug 16.

Wang E, Wang J, Zhu JX, Hao W, Wang L, Li Q, Zhang L, He W, Lu B, Lin H, Ma H, Zhang G, He Z (2008). Control of rice grain-filling and yield by a gene with a potential signature of domestication. Nature genetics 40:1370-1374.

Wang J, Nakazaki T, Chen S, Chen W, Saito H, Tsukiyama T, Okumoto Y, Xu Z, Tanisaka T (2009). Identification and characterization of the erect-pose panicle gene *Ep* conferring high grain yield in rice. (*Oryza sativa* L.). Theoretical and applied genetics 119:85-91.

William WW, Karol M, Piotr C (1997). Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity. Biotechniques 22(3):474-481.

Zhou Y, Zhu J, Li Z, Yi C, Liu J, Zhang H, Tang S, Gu M, Liang G (2009). Deletion in a quantitative trait gene QPE9-1 associated with panicle erectness improves plant architecture during rice domestication. Genetics 183:31.