Published Online First: August, 2024
https://doi.org/10.21123/bsj.2024.10073
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Diversity study of several domesticated rice (local cultivars) cultivated in the middle and south of Iraq using NGS technology

Raffal Ali Saloom**, Ali Mohammad Moner

Genetic Engineering and Biotechnology Institute for Postgraduate Studies, University of Baghdad, Baghdad, Iraq.
*Corresponding Author

Received 29/10/2023, Revised 02/01/2024, Accepted 04/01/2024, Published Online First 20/08/2024

Abstract

Due to the importance of the rice crop in Iraq, this study was conducted to determine the origin of the major varieties and understand the evolutionary relationships between Iraqi rice varieties and other Asian rice accessions that could be significant in the improvement of this crop. Phylogenomic analysis has been applied to clarify the relationship among rice species. Five varieties of Oryza sativa were obtained from the Agricultural Research Department in Iraq, (Amber33, Dijla, Ghadir, Baraka, and Black rice), and the whole genomic DNA was sequenced utilizing Next Generation Sequencing platforms based on DNA nanoball (DNB) technology. Sequences of 26 rice species were obtained from the NCBI Organelle Genome Resources database. Phylogenetic analysis of chloroplasts showed that they were separated into clades according to their region. Iraqi cultivars have been divided into two groups. The first one contains Amber33 and japonica NC_001320, while the other clade contains the Dijla, Ghadir, Baraka, and Black rice and indica NC_008155.

Keywords: chloroplast genome, evolutionary relationships, Oryza AA genome, phylogenetic analysis, rice (Oryza sativa).

Introduction

The rice plant, *Oryza sativa* L. belongs to the *Poaceae* family [1,2], and is one of the important grain crops in Iraq [3]. It has great nutritional value as a source of energy, protein, and carbohydrates[3,4]. It comes in second place after the wheat crop in terms of its economic importance and its role in food security in Iraq [5,6]. Chloroplast has a unique genome structure: mainly a large single copy, small single copy, inverted repeat A, and inverted repeat B this structure makes it difficult to assemble especially with short reads, because it could span multiple regions. The 135Kb chloroplast genome is maternally inherited and more highly conserved, making it a useful tool for evolutionary study and providing useful markers for phylogenetic investigations, the low complexity and high copy number of organelle genomes greatly facilitate their characterization [7,8]. With the advantages of next-generation sequencing, chloroplast genome sequences have been increasing dramatically during the last few years [9,10] . The whole genomic sequences of chloroplast plant breeders can more effectively comprehend the evolutionary links between accessions by knowing diversity patterns [11]. Studying the entire chloroplast genome sequence can provide comprehensive insight into the relationship with other *Oryza* species both wild and domesticated, unlike previous studies that focused on specific regions or some genes which did not represent the whole genetic materials [12-14]. Knowledge of diversity patterns allows plant breeders to better understand the evolutionary relationships among

Published Online First: August, 2024
https://doi.org/10.21123/bsj.2024.10073
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

accessions[2,15,16]. Rice chloroplast genome sequences provided an important tool for estimating genetic distance and determining evolutionary relationships among rice accessions, and also provided further information on the relationships between the studied varieties[9,17]. Therefore, this rice chloroplast-based study, aimed to provide more evidence about the domestication origin of Asian rice through the whole chloroplast genome sequences

Materials and Methods

Plant materials

Five *Oryza sativa* varieties were provided by the Office of Agricultural Research, Ministry of Agriculture, Baghdad, IRAQ Fig. 1 and Table 1.

DNA extraction and library preparation

Total genomic DNA was extracted from the individual seedling using CTAB protocol [18], sample concentration achieved the quantity and quality requirements of the library. According to the BGI procedure, 1µg genomic DNA was randomly fragmented by Covaris using microTUBE-15 to generate fragment sizes between 150-550bp.The fragmented genomic DNA was selected by the Agencourt AMPure XP-Medium kit. The average size of 200-400bp has been selected, and 150 paired-end read technique has been used for sequencing.

Figure 1. Differences in grains length and color of *Oryza sativa* used in this study.

Table 1. The Iraqi varieties used in this research.

Varieties	Grain characters	Days	Plant length	Production rate
Amber33	Aromatic, medium grain type	145 days	Long (150cm)	800 Kg/dunum
Black rice	Medium grain type	135 days	Long (145cm)	1350 Kg/dunum
Dijla	Medium grain type	130 days	Medium (100cm)	1800 Kg/dunum
Ghadir	Long-grain type	130 days	Medium (93cm)	1750 Kg/dunum
Baraka	Aromatic, Long-grain type	135 days	Medium (90cm)	1350 Kg/dunum

Data processing

The raw read data of five Iraqi varieties was subjected to quality control (QC) analysis using the Fastqc tool, to verify the quality of the data and determine the appropriate trimming score. The low-quality reads were trimmed to a minimum PHRED score of 30 using the "BB duck" tool[19].

Chloroplast genome assembly and SNP variant call and non-silent SNPs

A chloroplast genome of the Iraqi rice varieties was assembled by map to reference and denovo protocols[20]. The trimmed reads were mapped against the reference NC_001320, using a bowtie2 [21] Version 2.3 tool embedded in Geneious prime software. The following settings were used in alignment type: end to end and highly sensitive preset [7]. Variant call of polymorphism and amino acid alteration has been done utilizing Geneious SNP embedded tool for comparing local varieties to the referencing genome NC_001320.

Phylogenetic analysis

Chloroplast sequences of 26 Oryza species have been obtained from the NCBI Organelle Genome Resources database and was used in the Phylogenetic tree to compare with five Iraqi cultivars (Amber33, Black rice, Dijla, Ghadir and Baraka). The consensus chloroplast sequences of the Iraqi rice and the other domesticated rice accessions were used to perform a phylogenetic analysis using the Geneious software. The multiple alignments were conducted using the plugin MAFFT [22] Alignment. All sequences obtained were aligned using MAFFT tools (Auto, 1PAM/K=2 scoring matrix, 1.53 open gap penalty, and 0.123 offset value) with default parameters; subsequently, to analyze evolutionary relationships; phylogenetic

Published Online First: August, 2024
https://doi.org/10.21123/bsj.2024.10073
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

analysis with two different software packages was used: MrBayes v.3.2.6 [23](GTR, Gamma and100 bootstrapping). PHYML v.3.3 [24] (GTR, and100 bootstrapping) for more accuracy and high confidence in this analysis. Table 2 shows the Chloroplast sequences of 26 *Oryza* species obtained from the NCBI Organelle Genome Resources database.

Chloroplast genome draw

Organellar Genome DRAW (OGDRAW), an online tool was applied to each of the five chloroplast sequences (IBL= Black rice, IDJ= Dijla, IBRQ= Baraka, IGA= Ghadir, IAN33= Amber33), to draw graphical maps of plastid and mitochondrial genome annotations as well as their display as physical maps [25].

Table 2. Chloroplast sequences of 26 *Oryza* species obtained from the NCBI Organelle Genome Resources database.

Organism Name	Genome type	Size (Mb)	GC%	Chloroplast accession number	CDS	Release date
O. alta	CCDD	0.135175	38.9961	NC_034760	87	2017-05-24
O. australiensis	EE	0.135224	38.9487	NC_024608	83	2014-07-29
O. brachyantha	FF	0.134604	38.981	NC_030596	83	2016-07-12
O. eichingeri	CC	0.134821	39.0021	NC_034759	87	2017-05-24
O. glumipatula	AA	0.134583	38.9886	NC_027461	83	2015-07-14
O. grandiglumis	CCDD	0.13515	38.9937	NC_034761	87	2017-05-24
O. latifolia	CCDD	0.13519	38.9933	NC_034762	87	2017-05-24
O. longiglumis	-	0.135641	38.93	NC_034763	87	2017-05-24
O. longistaminata	AA	0.134567	38.9895	NC_027462	83	2015-07-14
O. malampuzhaensis	-	0.134643	38.9682	NC_053278	87	2021-03-09
O. meridionalis	AA	0.134558	39.0085	NC_016927	75	2012-02-28
O. meyeriana	-	0.136133	38.9443	NC_034765	86	2017-05-24
O. minuta	BBCC	0.135094	38.9647	NC_030298	89	2016-06-10
O. neocaledonica	-	0.13595	38.9503	NC_053276	87	2021-03-09
O. nivara SL10	AA	0.134494	39.0084	NC_005973	119	2004-07-12
O. officinalis	CC	0.134911	38.9983	NC_027463	83	2015-07-14
O. punctata	BB	0.134604	38.9743	NC_027676	100	2015-08-04
O. rhizomatis	CC	0.134796	39.0101	NC_034758	87	2017-05-24
O. ridleyi	HHJJ	0.135731	38.9174	NC_034764	87	2017-05-24
O. rufipogon	AA	0.134544	39.0029	NC_017835	77	2012-05-09
O. schlechteri	HHKK	0.135278	38.9516	NC_053277	86	2021-03-09
O. barthii	AA	0.134674	38.9897	NC_027460	82	2015-07-14
O. sativa	AA	0.134502	38.9979	NC_031333	100	2016-10-05
O. sativa Indica Group	AA	0.134496	38.9989	NC_008155	64	2006-06-16
O. sativa Japonica Group	AA	0.134525	43.5525	NC_001320	108	2009-04-15
O. glaberrima	AA	0.132629	38.96	NC_024175	83	2014-06-09

Results and Discussion

The sequencing of the five Iraqi varieties generated about 85.7 Gb of data containing 342 million 150-bp paired-end reads. The number of reads per sample ranged from 69.1 to 66.9 million paired-end reads. These data can provide an estimated average coverage of 27 to 28X of the whole rice genome, which is more than enough for chloroplast genome assembly. When raw data were trimmed to the quality limit of 30 percent, the number of reads decreased to almost 50% but the quality was raised to approximately 20 times Table 3, although there was a big reduction in the number of reads. It is still enough to assemble a good quality assembly [26].

Published Online First: August, 2024
https://doi.org/10.21123/bsj.2024.10073
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Table 3. Differences in the number of readings and the expected errors before and after trimming.

Cultivars	Number of read sequences before trimmed	Expected errors before trimmed	Number of read sequences after trimmed	Expected errors after trimmed
Amber33	67.558.264	38.332.906	39.589.214	1.671.545
Black rice	69.105.230	37.590.940	41.104.072	1.729.126
Dijla	69.203.434	37.688.941	40.918.886	1.731.574
Ghadir	69.553.738	43.561.691	37.867.516	1.626.384
Baraka	66.955.440	34.178.530	41.682.530	1.735.695

Chloroplast genome assembly

All read sequences have been mapped to the reference, *O. Sativa* NC-001320. Dual pipeline [20] with a good amount of reads (coverage) helped and improve the quality of the assembly and reduced the errors especially in the boundary regions of the IRA, IRB and LSC, SSC that contain similar sequences and confuse the assembler. Table 4 summarizes the IRA, IRB, LSC and SSC start and end positions in the local varieties [25] .

Table 4. Number of CDS, GC percentages, start and end of IRA, IRB, LSC, and SSC of local studied varieties.

Cultivar/ accession number	CDS	GC%	region	from	to	length
Oryza-sativa-IAN33/ OR002139	107	39.00%	IRA	113749	134544	20795
			IRB	80606	101401	20795
			LSC	1	80606	80605
			SCC	101401	134544	33143
Oryza-sativa-IGA/ OR002143	107	39.00%	IRA	113700	134495	20795
			IRB	80557	101352	20795
			LSC	1	80557	80556
			SCC	101352	134495	33143
Oryza-sativa-IBRQ/ OR002141	107	39.00%	IRA	113700	134495	20795
			IRB	80557	101352	20795
			LSC	1	80557	80556
			SCC	101352	134495	33143
Oryza-sativa-IBL/ OR002140	107	39.00%	IRA	113699	134494	20795
			IRB	80556	101351	20795
			LSC	1	80556	80555
			SCC	101351	134494	33143
Oryza-sativa-IDJ/ OR002142	107	39.00%	IRA	113699	134494	20795
			IRB	80556	101351	20795
			LSC	1	80556	80555
			SCC	101351	134494	33143

Variant call of chloroplast genome

Variant calling has been taken after aligning five cultivar sequences (Amber33, Black rice, Dijla, Ghadir, Baraka) to a reference genome NC_001320 to identify the variation from the reference genome Table 5. It has been found that the total number of total that SNPs for each variety is (Amber33 =327, Black rice =503, Dijla =529, Ghadir =116, Baraka =524) and 45 non silent SNPs in all cultivars. The photosynthetic in angiosperms, the gene order, content, and rate of sequence evolution of protein-coding genes of the chloroplast genomes are generally conserved [27] . Those facts agree with our assemblies but there were several expected changes, as in Table 5; because of the geographical isolation from the original species and the local breeding programs that have been subjected to these varieties. This polymorphism had an impact on amino acid translation in fifteen genes, as shown in Table 6.

The most common non-silent polymorphisms (FNPs) in chloroplast genes were *psaA*, *psbB*, and

Published Online First: August, 2024
https://doi.org/10.21123/bsj.2024.10073
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

clpP with 7, 6 and 4 times respectively, in which changing amino acids could impact on these genes expression or splicing or either up or down regulate producing immature proteins that eventually affect the entire chlorophyll metabolism, electron transport and carbon assimilation. Photosystem and energy related genes had the highest number of alterations compared with other genes. It could be probably because of the high exposure to free electrons that are generate from sun radiation (especially UV) captures and transfers the energy. These changes might reflect positively or negatively on plant growth in general and its response to the biotic and abiotic environmental stress [28,29], but more investigation is needed to verify the importance of those FNPs

.

Table 5. Variant call of five chloroplast sequences against reference genome NC_001320, SNPs positions, amino acid alteration and genes names.

Position	Reference / variant	Amino acid change	Polymorphism type	Gene	Product	Variant sequences
3098	C -> T	G -> R	SNP (transition)	matK	maturase K	IAN33
8622	T -> C	S -> P	SNP (transition)	-	hypothetical protein	IAN33, IBL, IBRQ, IDJ, IGA
14231	A -> G	V -> A	SNP (transition)	-	hypothetical protein	IAN33, IBL, IBRQ, IDJ, IGA
27517	CG -> GC	FE -> LQ	Substitution	rpoC2	RNA polymerase beta" subunit	IAN33, IBL, IBRQ, IDJ, IGA
28019	G -> T	W -> L	SNP (transversion)	rpoC2	RNA polymerase beta" subunit	IBL, IBRQ, IDJ, IGA
29113	A -> G	N -> D	SNP (transition)	rpoC2	RNA polymerase beta" subunit	IBL, IBRQ, IDJ, IGA
39772	GC -> CG	RL -> SV	Substitution	psaA	photosystem I P700 chlorophyll a apoprotein A1	IAN33, IBL, IBRQ, IDJ, IGA
40251	G -> C	R -> G	SNP (transversion)	psaA	photosystem I P700 chlorophyll a apoprotein A1	IAN33, IBL, IBRQ, IDJ, IGA
40482	G -> C	R -> G	SNP (transversion)	psaA	photosystem I P700 chlorophyll a apoprotein A1	IAN33, IBL, IBRQ, IDJ, IGA
40684	A -> T	H -> Q	SNP (transversion)	psaA	photosystem I P700 chlorophyll a apoprotein A1	IAN33, IBL, IBRQ, IDJ, IGA
40689	ACT -> GAA	S -> F	Substitution	psaA	photosystem I P700 chlorophyll a apoprotein A1	IAN33, IBL, IBRQ, IDJ, IGA
40839	A -> T	S -> T	SNP (transversion)	psaA	photosystem I P700 chlorophyll a apoprotein A1	IAN33, IBL, IBRQ, IDJ, IGA
41145	G -> T	L -> I	SNP (transversion)	psaA	photosystem I P700 chlorophyll a apoprotein A1	IAN33, IBL, IBRQ, IDJ, IGA
41921	GC -> CG	G -> A	Substitution	ycf3	photosystem I assembly protein Ycf3	IAN33, IBL, IBRQ, IDJ, IGA
42896	C -> A	S -> I	SNP (transversion)	ycf3	photosystem I assembly protein Ycf3	IAN33, IBL, IBRQ, IDJ, IGA

Published Online First: August, 2024
https://doi.org/10.21123/bsj.2024.10073
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

49212	C -> G	R -> T	SNP (transversion)	ndhK	NADH dehydrogenase subunit K	IAN33, IBL, IBRQ, IDJ, IGA
53201	C -> G	R -> P	SNP (transversion)	atpB	ATP synthase CF1 beta subunit	IAN33, IBL, IBRQ, IDJ, IGA
64660	CG -> GC	VV -> VL	Substitution	-	photosystem I subunit IX	IAN33, IBL, IBRQ, IDJ, IGA
64689	CG -> GC	R -> A	Substitution	-	photosystem I subunit IX	IAN33, IBL, IBRQ, IDJ, IGA
66104	C -> A	T -> N	SNP (transversion)	rps18	ribosomal protein S18	IAN33, IBL, IBRQ, IDJ, IGA
66402	A -> G	S -> P	SNP (transition)	rpl20	ribosomal protein L20	IBL, IBRQ, IDJ, IGA
67982	G -> C	P -> A	SNP (transversion)	clpP	ATP-dependent Clp protease proteolytic subunit	IAN33, IBL, IBRQ, IDJ, IGA
68008	G -> A	T -> I	SNP (transition)	clpP	ATP-dependent Clp protease proteolytic subunit	IAN33, IBL, IBRQ, IDJ, IGA
68021	T -> C	N -> D	SNP (transition)	clpP	ATP-dependent Clp protease proteolytic subunit	IAN33, IBL, IBRQ, IDJ, IGA
68110	C -> A	S -> I	SNP (transversion)	clpP	ATP-dependent Clp protease proteolytic subunit	IAN33, IBL, IBRQ, IDJ, IGA
69349	C -> T	A -> V	SNP (transition)	psbB	photosystem II 47 kDa protein	IBL, IBRQ, IDJ, IGA
70225	GC -> CG	A -> R	Substitution	psbB	photosystem II 47 kDa protein	IAN33, IBL, IBRQ, IDJ, IGA
70278	G -> A	A -> T	SNP (transition)	psbB	photosystem II 47 kDa protein	IAN33, IBL, IBRQ, IDJ, IGA
70281	A -> T	I -> F	SNP (transversion)	psbB	photosystem II 47 kDa protein	IAN33, IBL, IBRQ, IDJ, IGA
70292	CGT -> GTC	R -> V	Substitution	psbB	photosystem II 47 kDa protein	IAN33, IBL, IBRQ, IDJ, IGA
70307	AG -> GA	TG -> TR	Substitution	psbB	photosystem II 47 kDa protein	IAN33, IBL, IBRQ, IDJ, IGA
77794	CG -> AC	R -> V	Substitution	rpl16	ribosomal protein L16	IBL, IBRQ, IDJ, IGA
77794	CG -> GC	R -> A	Substitution	rpl16	ribosomal protein L16	IAN33
84654	CG -> GC	AE -> AQ	Substitution	-	hypothetical protein	IAN33, IBL, IBRQ, IDJ, IGA
90343	T -> G	M -> L	SNP (transversion)	-	hypothetical protein	IAN33, IBL, IBRQ, IDJ, IGA
102839	T -> C	T -> A	SNP (transition)	ndhF	NADH dehydrogenase subunit 5	IAN33, IBL, IBRQ, IDJ, IGA

Published Online First: August, 2024
https://doi.org/10.21123/bsj.2024.10073
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

105778	AAGC -> GCTT	LS -> LL	Substitution	ccsA	cytochrome c biogenesis protein	IAN33, IBL, IBRQ, IDJ, IGA
106801	G -> A	A -> V	SNP (transition)	ndhD	NADH dehydrogenase subunit 4	IAN33, IBL, IBRQ, IDJ, IGA
110510	-> ATAACT	-> SY	Insertion	ndhI	NADH dehydrogenase subunit I	IAN33, IBL, IBRQ, IDJ, IGA
124775	A -> C	M -> L	SNP (transversion)	-	hypothetical protein	IAN33, IBL, IBRQ, IDJ, IGA
130465	CG -> GC	AE -> AQ	Substitution	-	hypothetical protein	IAN33, IBL, IBRQ, IDJ, IGA

Table 6. Frequency of FNP snps in gene in chloroplast genome.

NO.	Gene names	Product	Number of FNP snp repetitions
1	psaA	photosystem I P700 chlorophyll an apoprotein A1	7
2	psbB	photosystem II 47 kDa protein	6
3	clpP	ATP-dependent Clp protease proteolytic subunit	4
4	rpoC2	RNA polymerase beta" subunit	3
5	rpl16	ribosomal protein L16	2
6	ycf3	photosystem I assembly protein Ycf3	2
7	rpl20	ribosomal protein L20	1
8	rps18	ribosomal protein S18	1
9	matK	maturase K	1
10	ndhK	NADH dehydrogenase subunit K	1
11	ndhI	NADH dehydrogenase subunit I	1
12	ndhD	NADH dehydrogenase subunit 4	1
13	ndhF	NADH dehydrogenase subunit 5	1
14	ccsA	cytochrome c biogenesis protein	1
15	atpB	ATP synthase CF1 beta subunit	1

Phylogenetic analysis of the chloroplast genome

Two phylogenetic approaches, MrBayes and PhyML were used to analyze the 26 chloroplast genomes representing the *Oryza* species and five Iraqi cultivars (IBL= Black rice, IDJ= Dijla, IBRQ= Baraka, IGA= Ghadir, IAN33= Amber33). Although the results of the two phylogenetic methods showed the same topology, there were minor alterations at the end of some sub-clades [7]. The phylogenetic analysis showed that all species were grouped according to geographical origin, and the Iraqi cultivars were grouped with domesticated rice species (specifically with Asian rice). However, IAN33 clustered with the japonica cultivars while, the other clustered with indica cultivars, as shown in Fig. 2. This output enhanced our current understanding of the phylogenetic relationships of the AA-genome Oryza species from different continents and proved that all local Iraqi varieties were originally purely derived from Asian rice and did not contaminate with African or South American rice throughout a breeding program in the last several decades. The genetic distance among species followed by the ecotype clade distribution in the main clades and sub-clades Table 7, which agrees with historical facts about the origin of Iraqi rice varieties [30,31].

Chloroplast genome draw

The complete chloroplast sequences of five local varieties have been drawn. The genome size is 134,550 bp, which is similar to the already reported cp genome sizes of related *Oryza* species and is within the range of other angiosperms [17]. The chloroplast genome possessed a typical quadripartite structure, which includes a pair of inverted repeats (IRa and IRb) and separate SSC and LSC regions including protein-coding genes, tRNA genes, and rRNA genes Fig. 3. Genes drawn inside the circle were transcribed clockwise, and those outsides were

Published Online First: August, 2024
https://doi.org/10.21123/bsj.2024.10073
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

transcribed counter clockwise. Genes belonging to different functional groups were color coded. The darker gray color in the inner circle corresponded to the GC content, and the lighter gray color corresponded to the AT content [31].

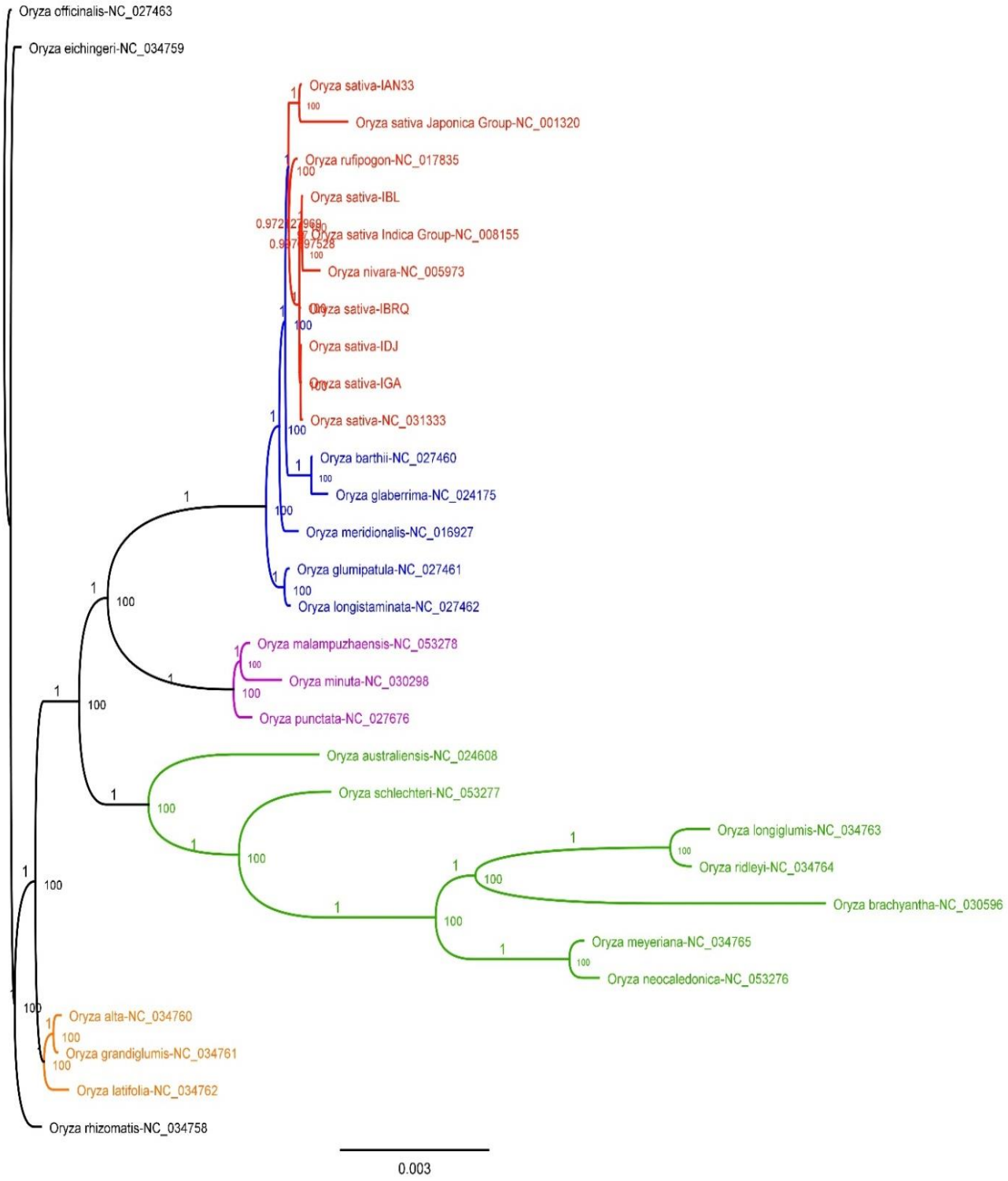

Figure 2. Chloroplast phylogenomic using PhyML and MrBayes software both have the same topology of 26 Oryza species from ncbi with five cultivars. *(IBL= Black rice, IDJ= Dijla, IBRQ= Baraka, IGA= Ghadir, IAN33= Amber33)

Published Online First: August, 2024
https://doi.org/10.21123/bsj.2024.10073
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

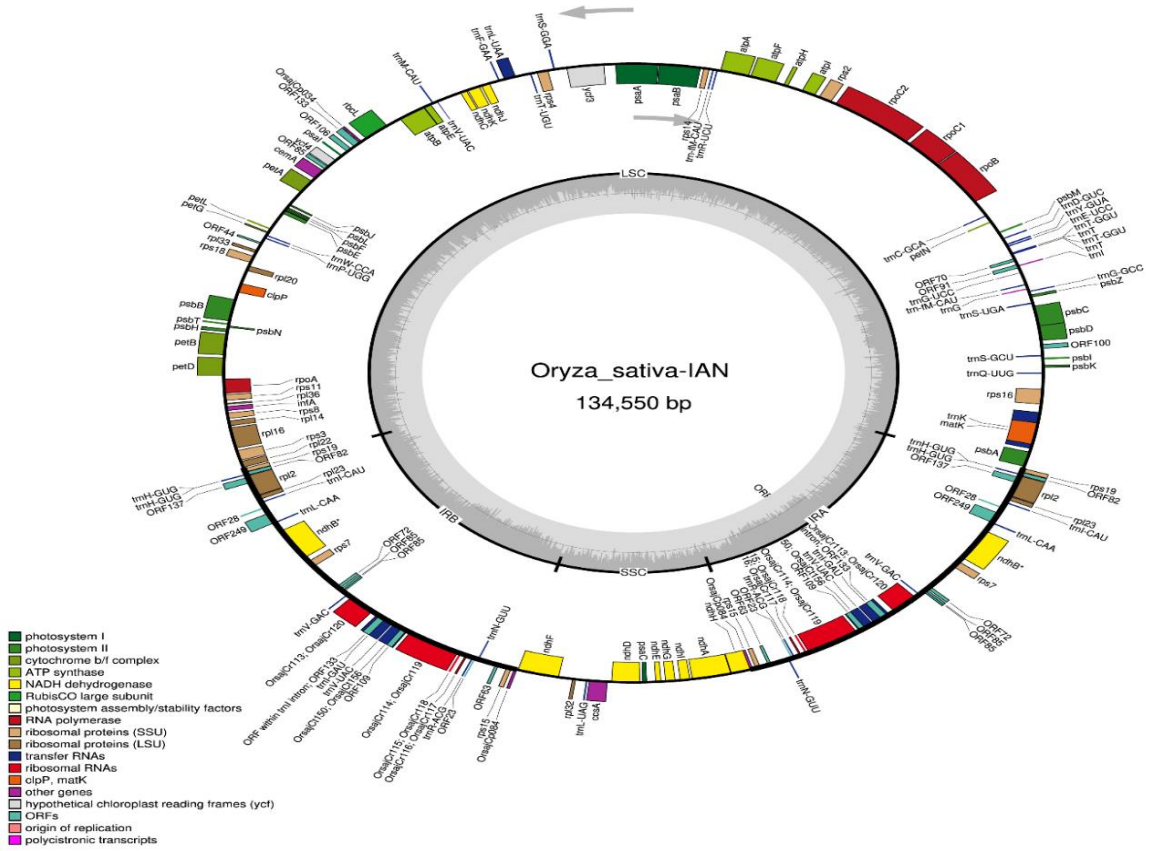

Figure 3. Chloroplast genome drawing of Amber33 variety using an Organellar Genome DRAW (OGDRAW) online tool

Conclusion

It has been concluded that the Chloroplast genomes of *Oryza sativa* Iraqi cultivars were grouped with other Asian domesticated rice and divided into two separate groups. There were 33 FNPs in the chloroplast genome of five Iraqi cultivars in 15 different genes and the *psaA, psbB,clpP* genes had the highest frequency. There are several variations in the chloroplast genome and genes especially Frequency functional nucleotide polymorphism, altering in amino acids might change the protein function or gene expression. The total SNPs for each cultivar were (Amber33 =327, Black rice =503, Dijla =529, Ghadir =116, Baraka =524) in the chloroplast genome.

Acknowledgement

The authors appreciate the cooperation of the Agricultural Research Department in Iraq in offering rice seeds.

Authors' Declaration

- Conflicts of Interest: None.
- We hereby confirm that all the figures and tables in the manuscript are ours. Furthermore, any figures and images that are not ours have been included with the necessary permission for re-publication, which is attached to the manuscript.
- The authors declare no competing financial interests.

Published Online First: August, 2024
https://doi.org/10.21123/bsj.2024.10073
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

- Ethical Clearance: The project was approved by the local ethical committee at the University of Baghdad.
- No animal studies are present in the manuscript.
- No human studies are present in the manuscript.
- No potentially identified images or data are present in the manuscript.

Authors' Contribution Statement

Both authors contributed to the article and approved the submitted version A.M designed the project, R.A. ran the experiment, A.M. and R.A. did the analysis and wrote the manuscript.

Supplemental Files

- Supplement 1.

References

1. Fouad AS, AlSobeai SM. In silico characterization of a cyclin dependent kinase-A (CDKA) and its coding gene in some Oryza species. Baghdad Sci J. 2020; 17(3): 760-771. http://dx.doi.org/10.21123/bsj.2020.17.3.0760
2. Fouad A S, Hafez R M. Molecular modeling and in silico characterization of a pathogenesis-related protein-10 (PR10) and its coding genes in some Oryza species. Baghdad Sci J. 2019; 16(4 (Suppl.)): 0993-1002. http://dx.doi.org/10.21123/bsj.2019.16.4(Suppl.).0993
3. Abakah AJS, Al-Mohammad MHS. Response of Yield Components of Some Black Rice Cultivars to Spraying with The Amino Acid Phenylalanin. IOP Conf Ser Earth Environ Sci. 2021. 910(1): 12020. http://doi.org/10.1088/1755-1315/910/1/012020
4. Civáň P, Craig H, Cox CJ, Brown TA. Three geographically separate domestications of Asian rice. Nat plants. 2015; 1(11): 1–5. http://dx.doi.org/10.1038/nplants.2015.164
5. Mohamed SJ, Salman OH. An economic Analysis of the Impact of the I Raqi Dinar Exchange Rate On the imported quantities of rice during the period 1990-2020. Iraqi J Agric Sci. 2022; 54(2): 542–52. https://doi.org/10.36103/ijas.v54i2.1730
6. Armanto ME. Improving rice yield and income of farmers by managing the soil organic carbon in South Sumatra landscape, Indonesia. Iraqi J Agric Sci. 2019; 50(2): 653–61. https://doi.org/10.36103/ijas.v2i50.665
7. SINGH, Bhupinder Pal, et al. CpGDB: a comprehensive database of chloroplast genomes. Bioinformation, 2020, 16.2: 171 https://doi.org/10.6026/97320630016171
8. Bogdanova Vera S. Genetic and molecular genetic basis of nuclear-plastid incompatibilities. Plants. 2019; 9(1):23-40 https://doi.org/10.3390/plants9010023
9. Cubry P, Tranchant-Dubreuil C, Thuillet A-C, Monat C, Ndjiondjop M-N, Labadie K, et al. The rise and fall of African rice cultivation revealed by analysis of 246 new genomes. Curr Biol. 2018; 28(14): 2274–2282. https://doi.org/10.1016/j.cub.2018.05.066
10. Rhoads A, Au KF. PacBio sequencing and its applications. Genom Proteom Bioinform. 2015; 13(5): 278–289. https://doi.org/10.1016/j.gpb.2015.08.002
11. Lateef AA, Garuba T, Abdulkareem KA, Olayinka BU, Olahan GS, Adeyemi SB, et al. Molecular Characterization of Potential Crop Pathogens Associated with Weeds as Endophytes in Uniilorin Plantations, Nigeria. Baghdad Sci J. 2022; 19(6): 1201-1211. https://dx.doi.org/10.21123/bsj.2022.5999
12. Moner AM, Furtado A, Henry RJ. Two divergent chloroplast genome sequence clades captured in the domesticated rice gene pool may have significance for rice production. BMC Plant Biol. 2020; 20(1): 1–9. https://doi.org/10.1186/s12870-020-02689-6
13. Cheng L, Nam J, Chu S H, Rungnapa P, Min, M H, Cao Y, et al. Signatures of differential selection in chloroplast genome between japonica and indica. Rice. 2019; 12, 1-13. https://doi.org/10.1186/s12284-019-0322-x
14. Al-Barhawee NIK, Ahmed JMY. Using Sequencing Technique for Diagnostic Different Species of Genus Rhizobium Which Isolated from Legume Plants. Iraqi J Sci. 2022; 63(10): 4213–4224. https://dx.doi.org/10.24996/ijs.2022.63.10.8
15. Gao L Z, Liu Y L, Zhang D, Li W, Gao J, Liu Y, et al. Evolution of Oryza chloroplast genomes promoted adaptation to diverse ecological habitats. Common

Published Online First: August, 2024
https://doi.org/10.21123/bsj.2024.10073
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Biol. 2019; 2(1): 278-291. https://doi.org/10.1038/s42003-019-0531-2
16. Jasim BN, Al-Salihy AA, Moner AM. The Partial DNA Sequencing and Phylogenic Analysis of Tomato yellow leaf curl virus Isolated from Iraqi tomato. Iraqi J Biotechnol. 2020; 19(1): 40–55.
17. Al-Hadeithi ZS, Al-Kazaz AKA, Al-Obaidi BK. Genetic diversity and relationships among Iraqi barley cultivars using RAPD–PCR technique. Iraqi J Agric Sci. 2012; 43:117–124.
18. Gitzendanner M A, Soltis P S, Wong G K S, Ruhfel BR, Soltis DE. Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. Am J Bot. 2018; 105(3): 291-301. https://doi.org/10.1002/ajb2.1048
19. Bushnell B, Rood J, Singer E. BBMerge – Accurate paired shotgun read merging via overlap. PLoS One. 2017; 12(10): 1–15. https://doi.org/10.1371/journal.pone.0185056
20. Saloom RA, Moner AM. Mapping to reference is an efficient approach to achieve sufficient consensus for phylogenomic studies "Oryza chloroplast genome as a case study." Iraqi J Biotechnol. 2022; 21(2). 439-445
21. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9(4): 357–359. https://doi.org/10.1038/nmeth.1923
22. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013 Apr; 30(4): 772–780. https://doi.org/10.1093/molbev/mst010
23. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinform. 2001 Aug; 17(8): 754–5. https://doi.org/10.1093/bioinformatics/17.8.754
24. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010; 59(3): 307–321. https://doi.org/10.1093/sysbio/syq010
25. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019; 47(W1): W59–64. https://doi.org/10.1093/nar/gkz238
26. Fujino K, Hirayama Y, Obara M, Tomohito Ikegaya. Introgression of the chromosomal region with the Pi-cd locus from Oryza meridionalis into O. sativa L. during rice domestication. Theor. Appl. Genet. 2019;132(7):1981–90. https://doi.org/10.1007/s00122-019-03332-1
27. ZHOU, Jiawu, et al. Interspecific Hybridization Is an Important Driving Force for Origin and Diversification of Asian Cultivated Rice Oryza sativa L. Front. Plant Sci, 2022, 13: 932737. https://doi.org/10.3389/fpls.2022.932737
28. Hoban S, Bruford M, Jackson J, Lopes-Fernandes M, Heuertz M, Hohenlohe, et al. Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. Biol Conserv. 2020; 248: 108654. https://doi.org/10.1016/j.biocon.2020.108654
29. Li Y, Yu C, Mo R, Zhu Z, Dong Z, Hu X, et al. Screening and Verification of Photosynthesis and Chloroplast-Related Genes in Mulberry by Comparative RNA-Seq and Virus-Induced Gene Silencing. Int J Mol Sci. 2022; 23(15): 8620-8638. https://doi.org/10.3390/ijms23158620
30. Li B, Zheng Y. Dynamic evolution and phylogenomic analysis of the chloroplast genome in Schisandraceae. Sci Rep. 2018; 8(1): 1–11. https://doi.org/10.1038/s41598-018-27453-7
31. Badro H, Furtado A, Henry R. Relationships between Iraqi rice varieties at the nuclear and plastid genome levels. Plants. 2019; 8(11): 481-495.

Published Online First: August, 2024
https://doi.org/10.21123/bsj.2024.10073
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

دراسة التنوع الوراثي لعدد من اصناف الرز العراقي المستزرع في وسط وجنوب العراق باستخدام تقنية تسلسل الجيل القادم

رفل علي سلوم، علي محمد منير

معهد الهندسة الوراثية والتقنيات الاحيائية للدراسات العليا، جامعة بغداد، بغداد، العراق

الخلاصة

نظرًا لأهمية محصول الأرز في العراق ، أجريت هذه الدراسة لتحديد أصول الأصناف الرئيسية وفهم العلاقات التطورية بين أصناف الأرز العراقية وغيرها من انواع الأرز الآسيوية التي يمكن أن تكون ذات اهمية كبيرة في تحسين هذا المحصول. تم الحصول على خمسة أنواع من قسم البحوث الزراعية في العراق, ومن بين هذه الاصناف (العنبر 33 ، دجلة ، غدير ، البركة ، والأرز الأسود) وقد تم تسلسل الحمض النووي الجيني باكملة باستخدام منصات تسلسل الجيل التالي القائمة على تقنية DNA nanoball (DNB),تم الحصول على تسلسل 26 نوعاً من قاعدة بيانات موارد جينوم NCBI .اظهر التحليل التطوري بنائاً على البلاستيدة الخضراء الانواع قسمت الى مجاميع وفقاً الى موقعهم الجغرافي تم تقسيم الأصناف العراقية إلى مجموعتين المجموعة الاولى تحتوي Amber33 و Japonica NC_001320 بينما تحتوي مجموعة ثانية على Dijla و Ghadir و Baraka و Black Rice و Indica NC_008155.

الكلمات المفتاحية: جينوم البلاستيدة الخضراء، العلاقات التطورية، أوريزا AA جينوم، تحليل شجرة التطور الوراثي، الأرز (أوريزا ساتيفا).