J.Natn.Sci.Foundation Sri Lanka 2017 45 (3): 229 - 237
DOI: http://dx.doi.org/10.4038/jnsfsr.v45i3.8187

RESEARCH ARTICLE

The molecular evolution of *Ghd7* orthologs is not in line with the speciation events in genus *Oryza*

Asanga D. Nagalla[1], Dimanthi V. Jayatilake[2], Venura Herath[2*] and Mangalika Wickramasinghe[2]
[1] *Postgraduate Institute of Agriculture, University of Peradeniya, Peradeniya.*
[2] *Department of Agricultural Biology, Faculty of Agriculture, University of Peradeniya, Peradeniya.*

Revised: 09 December 2016; Accepted: 19 January 2017

Abstract: Under long day conditions, the *grain number, plant height, and heading date7* (*Ghd7*) is a key regulatory gene in the rice specific flowering pathway, with significant pleotropic effects on yield related traits such as the number of spikelets per panicle, and heading date. The *Ghd7* codes for a CO, CO-LIKE and TIMING OF CAB1 (CCT) domain, which is unique from the other CCT family proteins. In the current study, the evolutionary relationships of putative orthologs of *Ghd7* within the genus *Oryza* is assessed *via* computational approaches. The relationships were deduced based on the nucleotide diversity, gene tree and synteny of *Ghd7* orthologs of nine *Oryza* species using genomic, coding, and polypeptide sequences of the putative *Ghd7* orthologs and adjacent genes retrieved from the Gramene database. The coding sequences (CDS) were found to be highly conserved (> 95 %) across the *Oryza* species, including the CCT domain. The length and the composition of the exons were not conserved between species. The exon-intron boundaries depicted an intron phase where codons in exons were not interrupted by introns. The reconstructed gene tree revealed earlier divergence of *Oryza brachyantha* (diploid F genome) and *Oryza punctata* (diploid B genome) from the diploid species carrying an A genome. The A genome carriers were grouped into two strongly supported monophyletic clades, with clear separation of the Asian-origin species from the African-origin species. The movement of the syntenic blocks comprising putative orthologs of *Ghd7* and its adjacent genes in 0.5 Mbp region revealed complex relationships, especially in ancestral species and species with African descent as the 1:1 putative orthologs were found in non-predictor chromosomes. Hence, the evolutionary pathway of *Ghd7* was found to be complex and was governed by geographic origins, and gene movement across syntenic and non-syntenic regions.

Keywords: Comparative genomics, gene tree, *Ghd7*, *Oryza*, synteny.

INTRODUCTION

Rice (*Oryza sativa* L.) is the second largest cereal crop produced in the world with an estimated annual production of 494.9 million tons in the year 2014/2015 (FAO, 2016). It is the food staple for more than half of the world's population, of which more than 90 % are Asians (Mohanty, 2013). The current world population is nearly 7.3 billion and considering the period 2000 − 2010, the population growth rate was approximately 3.5 % (FAO, 2016). Accordingly, the food production needs to be increased to match the growing demands, especially in staple food crops such as rice.

Rice yield is determined by the number of panicles per plant, number of spikelets per panicle and 1000-grain weight. The flowering time has a major effect on yield parameters and is under the regulation of many genes including *HEADING DATE 3a* (*Hd3a*) and *RICE FLOWERING LOCUS T 1* (*RFT1*) (Komiya *et al*., 2008). The *GRAIN NUMBER*, *PLANT HEIGHT*, *AND HEADING DATE 7* (*Ghd7*) regulates genes involved in flowering as a response to photoperiodism. The *Ghd7* alters the expression of the *EARLY HEADING DATE 1* (*Ehd1*) through the *CONSTANS-LIKE 10* (*OsCOL10*) (Tan *et al*., 2016), which in turn regulates *Hd3a* and *RFT1*, two major flowering genes in rice (Itoh & Izawa, 2013; Shrestha *et al*., 2014). Further, the *Ghd7* is known to interact with the expression of genes delivering drought tolerance, such as genes involved in scavenging/ producing reactive oxygen species (ROS) and those involved in abiotic stress-responses (Weng *et al*., 2014).

* Corresponding author (venurah@pdn.ac.lk)

Under drought the expression of *Ghd7* is known to get suppressed (Weng *et al*., 2014) and thus, could affect the yield when subjected to adverse abiotic stress conditions. Hence, *Ghd7* is an important gene for floral transition, yield determination and abiotic stress responses in rice.

The reconstruction of a gene tree using orthologs from close relatives would reveal their evolutionary connections (Wang *et al*., 2009). Given the importance of *Ghd7* in flowering time regulation and drought response, knowing its evolutionary roots will shed knowledge on how it evolved. However, a gene tree may not necessarily reflect the exact pathway in which speciation occurred, especially if the gene sequence is shorter (Pamilo & Nei, 1988). An insight of the events that took place during the evolution of *Ghd7* can be understood based on the variations in gene structure and composition (regulatory elements, motifs, domains, and intron/exon regions) and by understanding the syntenic relationships in comparable genomes.

The *Ghd7* evolution pre-dates the divergence of rice from other grass species such as purple false brome (*Brachypodium distachyon*), maize (*Zea mays*) and sorghum (*Sorghum bicolor*) (Yang *et al*., 2012). Comparison of the *Ghd7* orthologous regions in *Oryza* species has revealed that there is high collinearity within the genus *Oryza* (Yang *et al*., 2012). However, *Ghd7* homologs of ancestors *Oryza barthii* and *Oryza longistaminata* were not included in the synteny map constructed by Yang *et al*. (2012). Further, a *Ghd7* gene tree exclusive to genus *Oryza* has not been reconstructed to reveal the evolution of the *Ghd7*.

Studying the *Ghd7* orthologs within the genus *Oryza* would shed insight about the recent evolutionary events and gene movements within the genus and may reveal a direct line of gene descent that may otherwise be confounded at the species level. Hence, in the current study, the DNA polymorphisms and nucleotide diversity of the *Ghd7* putative orthologs in genus *Oryza* were compared, a gene tree was reconstructed and the syntenic relationships of the chromosomal regions harbouring *Ghd7* were examined to understand the evolutionary trends of the *Ghd7* in the genus *Oryza*.

METHODOLOGY

Nucleotide diversity analysis and reconstruction of *Ghd7* gene tree

The genomic sequence of *Ghd7* of *O. sativa* ssp. *japonica* (OS07G0261200) was retrieved from the Gramene database (Monaco *et al*., 2014, available at *http://www.gramene.org*) and was used as a query to retrieve genomic DNA sequences of its 1:1 putative orthologs belonging to the genus *Oryza* and its sister-genus *Leersia*. Multiple sequence alignment of the genus *Oryza* sequences were carried out in Geneious v7.1 (Drummond *et al*., 2011) using ClustalW (Thompson *et al*., 1994) platform with a gap open cost of 15 and a gap extend cost of 6.66.

The DNA sequence variation was assessed by calculating the nucleotide diversity (Pi) (Jukes & Cantor, 1969) in DnaSP v5.10 (Librado & Rozas, 2009). The species were categorised according to their origins (as Asian-origin and African-origin). Considering all species and categories, the Pi values were calculated for the genomic region and for the coding sequence (CDS).

A rooted gene tree was reconstructed with the *Ghd7* putative ortholog of *Leersia perrieri* (LPERR07G08310) as the outgroup in RAxML v7.2.6 (Stamatakis, 2006) using an un-gapped sequence alignment of *Oryza* spp., excluding InDel regions and following a maximum likelihood method with GTRGAMMA substitution model with 200 independent searchers. The branch support was estimated using 1000 bootstrap (BS) pseudoreplicates.

Identification of structural components in *Ghd7* orthologs

The sequences of the 5' and 3' untranslated regions, exons, introns of the *Ghd7* (OS07G0261200) and its putative orthologs of genus *Oryza* (species *O. sativa* ssp. *japonica*, *O. sativa* ssp. *indica*, *O. barthii*, *O. punctata*, *Oryza glaberrima*, *Oryza glumaepatula*, *O. longistaminata*, *Oryza brachyantha*, *Oryza nivara*, and *Oryza rufipogon*) were retrieved from the Plant Genes 49 database using the Gramene embedded BioMart tool (Haider *et al*., 2009, available at *http://ensembl.gramene.org/biomart/*). Further, the CCT domain (CO, CO-LIKE and TIMING OF CAB1) of the polypeptide sequence (UniProt ID: E5RQA1) of *O. sativa* ssp. *japonica* GHD7 was retrieved from the UniProt database (The UniProt Consortium, 2014, available at *http://www.uniprot.org*/). Of the retrieved sequences, *O. longistaminata* was excluded from analysis as the sequence was highly ambiguous. Intron phase distribution analysis was conducted based on the method described by Long *et al*. (1995) with three classes: phase 0 (intron is located between two codons), phase 1 and phase 2 (intron is placed within a codon, after the first and second nucleotide, respectively). The structural components and the CCT domains of the *Ghd7* was annotated and was visualised using the software - Illustrator for the Presentation and Visualization of Biological Sequences (IBS) v1.0 (Liu *et al*., 2015).

Synteny mapping of *Ghd7*

The anchor position of the *Ghd7* was identified on the *O. sativa* ssp. *japonica* genomic assembly in Gramene database. The sequences of genes over a region of 0.5 Mbp spanning over the *Ghd7* were retrieved/ and used as queries in a BLASTn search in Gramene database to identify similar regions in the genomic assemblies of *O. barthii*, *O. glaberrima*, *O. glumaepatula*, *O. nivara*, *O. punctata*, *O. rufipogon* and *O. sativa* ssp. *indica.* The anchor positions were identified based on the best hit positions (E-value of 0.0; identity of > 94 %). The gene position and the order of the orthologs of *Ghd7* and its adjacent/overlapping genes were compared to that of *O. sativa* ssp. *japonica* in a synteny map constructed using IBS v1.0 (Liu *et al*., 2015).

RESULTS AND DISCUSSION

The *Ghd7* (OS07G0261200) is located on the short-arm of the chromosome 7 of *O. sativa* and is known to code for a single transcript producing a unique CCT domain family protein, important for floral transition involved in the regulation of plant height, heading date and the number of spikelets per panicle (Xue *et al*., 2008). In the current study, the *Ghd7* of *O. sativa* ssp. *japonica* and its putative orthologs belonging to genus *Oryza* were used to understand the gene structure, nucleotide diversity and the phylogenetic relationships of the *Ghd7*.

Exploring the genetic diversity of *Ghd7*

A Gramene database search of *O. sativa* ssp. *japonica Ghd7* (OS07G0261200; *OSGhd7*) revealed 28 putative orthologs (Table 1), out of which 12 were 1:1 orthologs. Of those, nine belonged to the genus *Oryza* and the remaining three were *L. perrieri* (*LPERRGhd7*), *Sorghum bicolor* (*SBGhd7*) and *Zea mays* (*GRMZMGhd7*). The putative orthologs belonged to the subfamilies Panicoideae (*S. bicolor* and *Z. mays*) and Oryzoideae (*Oryza* spp., and *L. perrieri*) implying that the origin of *Ghd7* gene would have pre-dated the divergence of these subfamilies from each other.

Table 1: Orthologs of *Oryza sativa* ssp. *japonica Ghd7* (*Os07g0261200*)

Species	Gramene database ID	Ortholog name	Target identity (%)*	Query identity (%)**
O. punctata	OPUNC07G08790	*OPUNCGhd7*	52	53
O. barthii	OBART07G06400	*OBARTGhd7*	97	97
O. sativa ssp. *indica*	BGIOSGA024502	*BGIOSGAGhd7*	94	94
O. nivara	ONIVA01G32960	*ONIVAGhd7*	88	98
O. glumaepatula	OGLUM01G41660	*OGLUMGhd7*	88	98
O. glaberrima	ORGLA07G0080500	*ORGLAGhd7*	87	97
O. brachyantha	OB07G16920	*OBGhd7*	45	19
O. rufipogon	ORUFI07G09550	*ORUFIGhd7*	44	53
Leersia perrieri	LPERR07G08310	*LPERRGhd7*	35	35
Sorghum bicolor	Sb06g000570	*SBGhd7*	27	26
Zea mays	GRMZM2G381691	*GRMZMGhd7*	22	20
Triticum aestivum	Traes_1AL_2EE300C85	-	42	21
Brassica oleracea	Bo9g163650	-	35	11
Populus trichocarpa	POPTR_0105s00200	-	35	16
Setaria italica	Si011920m.g	-	34	28
S. italica	Si039184m.g	-	31	30
Musa acuminata	GSMUA_Achr5G13470_001	-	30	21
Amborella trichopoda	AMTR_s00111p00022140	-	18	12
Selaginella moellendorffii	SELMODRAFT_141010	-	13	14
Ostreococcus lucimarinus	OSTLU_37312	-	10	17
Chlamydomonas reinhardtii	CHLREDRAFT_149159	-	5	20
Triticum aestivum	Traes_4BL_1FE4A71E6	-	44	15
T. aestivum	Traes_1BL_688EF6A1A	-	40	20
T. urartu	TRIUR3_25590	-	39	20
T. urartu	TRIUR3_08750	-	27	23
T. urartu	TRIUR3_08751	-	27	21
T. urartu	TRIUR3_30403	-	26	21

* Percentage of the target sequence matching the query sequence
** Percentage of the query sequence matching the target sequence

In species belonging to genus *Oryza*, the *Ghd7* consists of a variable number of predicted exons (one to three exons). In *O. sativa* ssp. *japonica*, the *Ghd7* consists of two exons (Figure 1) spread over a genic region of 3854 bp (position 9152377-9155030 in chromosome 7). Similarly, its putative orthologs (Table 1) in *O. sativa* ssp. *indica* (*BGIOSGAGhd7*), *O. punctata* (*OPUNCGhd7*) and *O. barthii* (*OBARTGhd7*) also consists of two exons. However, in contrasting, putative orthologs in *O. rufipogon* (*ORUFIGhd7*), *O. glaberrima* (*ORGLAGhd7*), *O. glumaepatula* (*OGLUMGhd7*), and *O. nivara* (*ONIVAGhd7*) contain three exons and *O. brachyantha* (*OBGhd7*) contains only one exon (Figure 1).

Pairwise identity of all aligned genomic sequences belonging to the 10 *Oryza* species was 79 % (data not shown), indicating that the *Ghd7* genomic sequence is generally conserved. However, the CDS are much conserved with a pairwise identity of 95.5 % (data not shown), indicating that most DNA polymorphisms were concentrated in the intron region. Interestingly, the CDS of species was not formed based on a ridged exon boundary, as between species the exons have a variable sequence composition and length. For example, the 1st exon of the putative ortholog *ORUFIGhd7* (189 bp in size) consists of a 64 bp sequence found in the exon 1 of *ORGLAGhd7*, *OGLUMGhd7* and *ONIVAGhd7,* and a 125 bp region from the intron 1 in *ONIVAGhd7* and *ORGLAGhd7*, which was also found in a region upstream to the translation start site of exon 1 in *BGIOSGAGhd7*. Further, part of the intron 1 sequence of *ORUFIGhd7* is included in the exon 1 or 2 in other species. Thus, a clear demarcation of exon boundaries was not seen across the species barrier at all times. Due to the variations in the number of exons and the length of exon sequences, the *Ghd7* CDS varied in length among species, with *O. rufipogon* carrying the longest CDS (933 bp). Despite the length and composition of the CDS, in all species the nucleotide sequence of the CCT domain was present and was in the last exon when there was more than one exon (Figure 1). However, all exon-intron boundaries were showing intron insertions between two codons and hence, with respect to two adjacent introns, only symmetric exons were detected in the studied *Ghd7* orthologs (intron phase 0, 0; indicating an intron phase 0 in both 5' intron and 3' intron of an exon, respectively). Such symmetric exons are indicative of a reading frame that is not disturbed due to the addition of the introns (Long *et al.*, 1995).

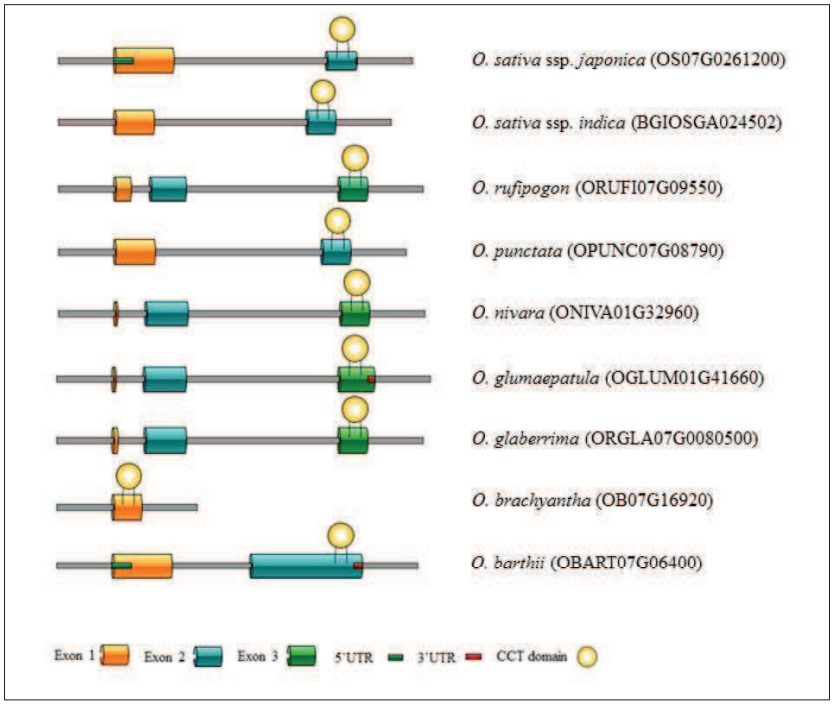

Figure 1: Gene structure of *Ghd7* genes in *Oryza* species. The figure illustrates the gene structure of the *Ghd7* gene orthologs in *Oryza* species and the positions of the CCT domains.

For *Ghd7*, a total of 12 haplotypes: nine from *O. sativa* ssp. *indica* (haplotypes 2, 3, 4, 6, 7, 8, 9, 10 and 11) and three from *O. sativa* ssp. *japonica* (haplotypes 0, 1, and 5) have been reported to-date (Lu *et al*., 2012). The *Ghd7* of *O. sativa* ssp. *japonica* cv. Nipponbare and *O. sativa* ssp. *indica* cv. 93-11 were similar to haplotypes 1 and 2, respectively. In previous studies it was reported that the wild relative *O. rufipogon* accessions resemble the haplotypes 1 or 3 [as haplotype 1 by Xue *et al.* (2008) in accession IRGC80752, and haplotypes 1 and 3 by Lu *et al.* (2012) in three unidentified accessions]. Interestingly, *O. rufipogon Ghd7* ortholog (*ORUFIGhd7*) mostly resembled the haplotype 1, with variations at a few SNPs and INDELs in the gene sequence. Apart from an assessment on the haplotype diversity in approximately 100 accessions of *O. sativa* (involving both *japonica* and *indica* subspecies) and few accessions of *O. rufipogon* (Xue *et al*., 2008; Yang *et al*., 2012), no studies have been conducted to assess the allelic diversity of the remaining *Oryza* species. Availability of sequences from different accessions within the species will enable the construction of a phylogeny of the *Ghd7* gene at species level, revealing further evolutionary trends. Hence, a comprehensive haplotype mining by sequencing multiple accessions of species belonging to the genus *Oryza,* representing wider geographic origin is recommended.

In *O. sativa* a total of eight predicted GHD7 proteins (GHD7-0, GHD7-1, GHD7-2, GHD7-3, GHD7-4, GHD7-5, GHD7-6 and GHD7-7) have been identified. Out of which the GHD7-0 predicted protein was identified as non-functional due to the absence of the CCT domain, as it was interrupted by a stop codon at the exon 1 (Xue *et al*., 2008; Lu *et al*., 2012; Yang *et al*., 2012). In addition, a *Ghd7* null-allele, commonly referred to as *Ghd7-0a* has also been reported. The species *O. punctata*, *O. barthii, O. rufipogon, O. glaberrima*, *O. glumaepatula*, *O. nivara* and *O. brachyantha* all carried a *Ghd7* putative ortholog in their genomes (*OPUNCGhd7*, *OBARTGhd7*, *ORUFIGhd7*, *ORGLAGhd7*, *OGLUMGhd7*, *ONIVAGhd7* and *OBGhd7*, respectively), hence, *Ghd7-0a* null allele was absent among the selected species. The peptide sequence of the predicted GHD7 proteins of the nine *Oryza* spp. does not have a premature stop codon. Thus, none produce the GHD7-0 predicted protein or any other protein that was a result of an early translation termination. Compared to *O. sativa*, the predicted peptide sequences of *O. rufipogon*, *O. punctata* and *O. brachyantha* were less conserved [< 95 % pairwise identity (data not shown)]. When the predicted peptide sequences of the remaining sequences of *O. barthii, O. glaberrima*, *O. glumaepatula*, and *O. nivara* were aligned to that of *O. sativa*, several non-synonymous mutations were detected (Table 2). In addition to the

Table 2: Non-synonymous substitutions in the GHD7 protein in selected *Oryza* species

Species	Position*	Amino acid change with respect to *O. sativa*	Polarity change	Acidity change
O. barthii	53	Serine to Cysteine	Polar to non-polar	Neutral-positive
O. barthii	68	Proline to Arginine	Non-polar to polar	Neutral-positive
O. barthii	114	Alanine to Proline	No	No
O. barthii	200	Alanine to Threonine	Non-polar to polar	No
O. barthii	218	Valine to Methionine	No	No
O. glaberrima	53	Serine to Cysteine	Polar to non-polar	Neutral-positive
O. glaberrima	68	Proline to Arginine	Non-polar to polar	Neutral-positive
O. glaberrima	114	Alanine to Proline	No	No
O. glaberrima	200	Alanine to Threonine	Non-polar to polar	No
O. glaberrima	218	Valine to Methionine	No	No
O. glumaepatula	8	Serine to Threonine	No	No
O. glumaepatula	53	Serine to Cysteine	Polar to non-polar	Neutral-positive
O. glumaepatula	200	Alanine to Threonine	Non-polar to polar	No
O. glumaepatula	226	Arginine to Lysine	No	No
O. nivara	166	Glycine to Serine	Non-polar to polar	No
O. nivara	199	Alanine to Proline	No	No
O. sativa indica	166	Glycine to Serine	Non-polar to polar	No
O. sativa indica	263	Proline to Alanine	No	No

* Transcription start site was considered as the starting position

changes that would affect the geometry of the predicted protein, some of these amino acid transitions could possibly alter the chemical properties of the protein such as polarity, charge, acidity and water affinity (Table 2). It was noted that the 43 amino acids marking the CCT domain is mostly conserved among the nine *Oryza* spp., except for a single amino acid change in *O. glumaepatula* (Arginine to Lysine) at the 7th amino acid position. However, this amino acid change in the CCT domain is expected to have a minimal impact on the function of the protein, as physical properties remained constant apart from the change in geometry.

The DNA polymorphism and nucleotide diversity of the *Ghd7* in genus *Oryza* were calculated using the genomic sequences and CDS, considering all nine *Ghd7* orthologs from genus *Oryza* and in groups made based on their origins (Asian and African). The putative orthologs *OSGhd7*, *BGIOSGAGhd7, ONIVAGhd7* and *ORUFIGhd7* were grouped as Asian-origin (Khush, 1997), and putative orthologs *ORGLAGhd7, OPUNCGhd7, OBGhd7* and *OBARTGhd7* were grouped as African-origin (Khush, 1997; Rakshit *et al*., 2007; Wang *et al*., 2014). The ortholog *OGLUMGhd7* of species *O. glumaepatula* found in South and Central America was grouped under African-origin based on the fact that it is closely related to *O. barthii* (Vaughan *et al*., 2003; Wang *et al*., 2014).

The Pi value reveals the average pairwise difference per site (Nei, 1987). Similar to the findings of Lu *et al.* (2012), the current study reports positive Pi values for the nucleotide diversity of *Ghd7*. Positive Pi values are indicative of a positive selection that occurred with respect to *Ghd7*. Given the pleotropic control of *Ghd7* on plant height, heading date and number of spikelets per panicle, a positive selection occurring at the *Ghd7* locus is a likely scenario. When all nine *Oryza* species (all-origins) were considered, the Pi value (average DNA polymorphism of all possible sequence comparisons; Jukes and Cantor, 1969) of the genomic sequences ($Pi_{genomic}$ = 0.24494) is higher compared to that of CDS (Pi_{CDS} = 0.05590). The lower Pi value of the CDS compared to the genomic sequence indicates a conserved CDS region, an attribute of a gene coding for a functional protein. Interestingly, within the Asian-origin group Pi_{CDS} (0.01908) was higher compared to $Pi_{genomic}$ (0.01015). This could be a result of the higher degree of polymorphism at the exon 1 of *O. rufipogon* (a fragment of 125 bp inserted to the 5' end of exon 1) compared to the rest. Further, the nucleotide diversity (both genomic and CDS sequence) was higher ($Pi_{genomic}$ = 0.50639; Pi_{CDS} = 0.09097) in African-origin, when compared to those with Asian-origin ($Pi_{genomic}$ = 0.01015; Pi_{CDS} = 0.01908). This indicates the parallel evolution of Asian and African-origin rice with independent isolated mutation events.

Mining evolutionary trends of *Ghd7*

The subtribe Oryzinae within the subfamily Oryzoideae, consists of genera *Leersia* and *Oryza*, which are expected to have diverged from each other 20 million years ago (MYA) (Khush & Brar, 2001). Thus, the reconstructed CDS phylogentic tree of *Ghd7* was rooted using the *Ghd7* putative ortholog of *L. perrieri* representing the sister-genus *Leersia*. The resulted phylogeny revealed strong evidence depicting evolutionary trends of *Ghd7* within the genus *Oryza* (Figure 2). According to Hillis and Bull (1993), a BS > 70 can be considered as strong evidence to consider a presented branching pattern as realistic at a 95 % confidence interval. In the CDS gene tree of *Ghd7*, *OBGhd7* is placed basally to the rest of the group, followed by *OPUNCGhd7* (Figure 2). The remaining species are separated into two strongly supported (BS = 100) monophyletic clades (Clade I and Clade II; Figure 2). Clade I consists of *OSGhd7, BGIOSGAGhd7, OLONGhd7, ORUFIGhd7* and *ONIVAGhd7*. Thus, this clade is generally represented by species with Asian-origin (Khush, 1997). Within Clade I the close relationship of *ORUFIGhd7* and *OSGhd7* is strongly supported (BS = 98), indicating the close relationship of *O. rufipogon* and *O. sativa*. However, the relationships of the remaining species cannot be expressed with confidence due to low bootstrap support. Clade II consists of putative orthologs of *ORGLAGhd7*, *OGLUMGhd7* and *OBARTGhd7* representing species with an African-origin (Khush, 1997; Khush & Brar, 2001). Even though the divergence of *OGLUMGhd7* from the other two orthologs *ORGLAGhd7* and *OBARTGhd7* is not strongly supported, the close relationship of *ORGLAGhd7* and *OBARTGhd7* is strongly supported (BS = 99). This is in agreement with Wang *et al*. (2014), where a close relationship of *O. barthii* and *O. glaberrima* has been reported. The putative orthologs *OBGhd7* and *OPUNCGhd7* are the first to diverge from the rest; both are of African-origin, and carry diploid BB and FF genomes, respectively (Khush & Brar, 2001). The remaining species in Clade I and II, all contain a diploid AA genome (Khush, 1997; Khush & Brar, 2001).

The wild rice *O. nivara* evolved from *O. rufipogon* and thus, these two species are the early ancestors of *O. sativa*, which is an Asian rice. The *indica* and *japonica* subspecies of *O. sativa* diverged 0.4 MYA and was domesticated from *O. nivara* (Ge *et al*., 2005).

The African rice *O. glaberrima* does not share a direct descent from *O. nivara* and *O. rufipogon*, hence evolved parallel to *O. sativa* (Khush & Brar, 2001). According to Wang *et al*. (2014), *O. glaberrima* was domesticated from *O. barthii* in West Africa nearly 3,000 years ago. The species *O. glumaepatula* is a South American rice species with close roots to *O. barthii*, thus, shares an ancestry with African-origin rice. Hence, the line of species evolution does not completely agree with the trends depicted by the *Ghd7* gene tree (Clade 1 and Clade II of Figure 2), however, some relationships are well supported. Nevertheless, it is not always necessary that a gene tree follow the trends of actual species evaluation, especially when a fewer number of gene homologs were used in the construction of the gene tree (Pamilo & Nei, 1988).

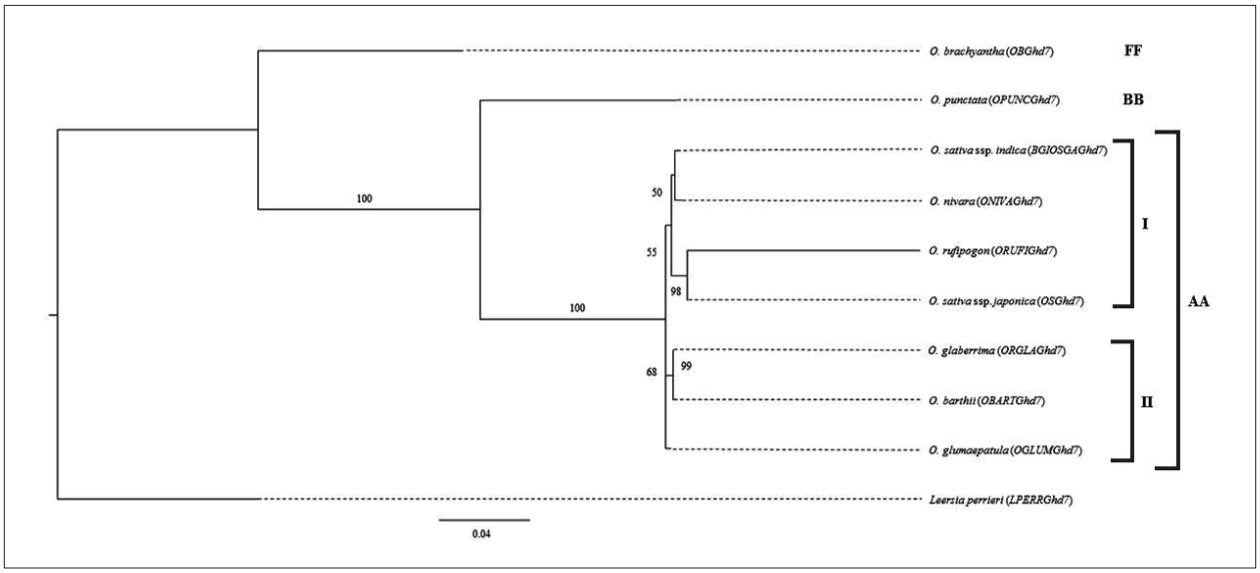

Figure 2: The *Ghd7* gene tree of genus *Oryza*. The letters AA, BB and FF represent the genomes, and I and II represent the monophyletic clades called at a bootstrap value of 100.

Changes in the gene structure and its chromosome position in closely related species provide an insight about its evolutionary relationships. Gene orthologs descend from a common ancestor and diverge, however, more often orthologs maintain the function of the ancestral gene (Koonin, 2005). The *Ghd7* has putative orthologs going beyond the genus *Oryza.* Similar to *OsGhd7* of *O. sativa* ssp. *japonica*, its putative orthologs *BGIOSGAGhd7*, *ORUFIGhd7*, *ORGLAGhd7*, *OBARTGhd7*, *OBGhd7* and *OLONGhd7* are mapped to the chromosome 7. However, the putative orthologs in *OGLUMGhd7* and *ONIVAGhd7* are mapped to chromosome 1 (Figure 3). Yang *et al.* (2012) reported a strong collinearity between the *Ghd7* orthologs in the genus *Oryza*, however, beyond the genus boundary, the synteny has become complex as a result of gene movement. The *Ghd7* putative orthologs were mapped to chromosome 3, 6 and 10 in *B. distachyon*, *S. bicolor* and *Z. mays*, respectively. These regions are in synteny with the chromosome 2 and 4 of *O. sativa*, and not with the chromosome 7 that carries the *Ghd7*.

The synteny map revealed similarities in the relative anchor positions of the *Ghd7* putative orthologs and the putative orthologs of the five adjacent/overlapping genes [two and three genes adjacent/overlapping to *Ghd7* (OS07G0261200) from proximal and distal sides spanning over a 0.5 Mbp region, respectively] in *O. sativa* ssp. *japonica*, *O. sativa* ssp. *indica*, *O. rufipogon* and in *O. glaberrima*. However, a much complex movement of chromosomal segments carrying homologous genomic regions were observed in *O. glumaepatula*, *O. punctata*, *O. nivara* and *O. barthii*, where the putative orthologs of *Ghd7* and adjacent/overlapping genes were mapped to a different location of the chromosome 7 itself (*O. barthii*) (Figure 3) or to an entirely different chromosome (to chromosome 1 or 12 in *O. glumaepatula*, *O. punctata* and *O. nivara*) (Figure 3).

The species evolution depicts the line of species descent of common Asian rice as *O. rufipogon* to *O. nivara* to *O. sativa.* However, the synteny blocks

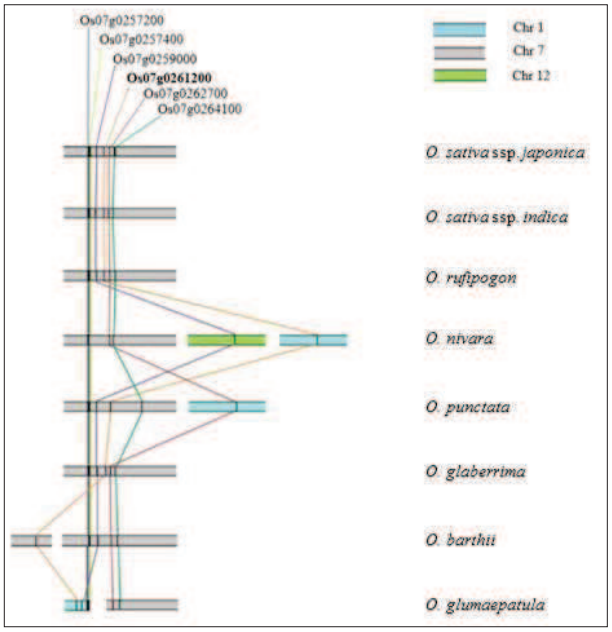

Figure 3: Synteny of putative orthologs of *Ghd7* and adjacent/ overlapping genes in the genus *Oryza*. The adjacent/ overlapping genes were selected from a 0.5 Mbp region of the *Oryza sativa* genomic sequence spanning over *Ghd7*.

carrying the putative orthologs of *Ghd7* and its adjacent/ overlapping genes indicate a complex gene movement. As an example, the *Ghd7* ortholog of *O. rufipogon* has moved from the chromosome 7 to chromosome 1 in *O. nivara*, and the putative ortholog of the gene adjacent to *Ghd7* in *O. rufipogon* (ORUFI07G09450) from chromosome 7 to chromosome 12 in *O. nivara* (Figure 3). The movements of these putative orthologs were reversed back again to chromosome 7 in *O. sativa*. This indicates a contradicting *Ghd7* gene evolution compared to the species evolution history of rice (Khush, 1997). Similarly, even though *O. barthii* and *O. glaberrima* are in a direct line of descent (Wang *et al*., 2014), the gene position of *Ghd7* shows movement within the chromosome 7. In the ancestor *O. barthii*, the gene was positioned away from the position where it is anchored in *O. glaberrima*. In addition, in *O. glumaepatula*, the *Ghd7* putative ortholog *OGLUMGhd7* and a block of putative orthologs with adjacent/overlapping genes (OS07G0257200, OS07G0257400 and OS07G0259000) were found to be on chromosome 1, instead of chromosome 7. In general, it is indicative that the putative *Ghd7* and adjacent/ overlapping gene orthologs of African-origin rice are subjected to a complex shuffling of positions compared to its Asian counterparts.

CONCLUSION

It is evident that *Ghd7* has undergone a positive selection in its evolutionary history and carries a comparatively conserved CDS region, especially at the CCT domain, the functional domain of the GHD7 protein. The *Ghd7* shows a complex evolutionary relationship that reflects the geographic separation of the Asian and African-origin rice, and their independent evolution. The syntenic relationships of the orthologs of *Ghd7* and its adjacent genes revealed a general collinearity within the *Oryza* genus, however, complex gene movements across non-syntenic genomic regions were revealed. Thus, the line of descent of *Ghd7* was not exactly in agreement with the relationships depicted by the species evolution.

REFERENCES

1. Drummond A.J. *et al*., (2011). *Geneious* v5.4. Available at: *http://www.geneious.com,* Accessed 15 September 2016.
2. Food and Agriculture Organization (FAO) (2016). *World Food Situation*. Available at: *http://www.fao.org/worldfoodsituation/csdb/en/*, Accessed 21 August 2016.
3. Ge S., Guo Y.L. & Zhu Q.H. (2005). Molecular phylogeny and divergence of the rice tribe (Oryzeae), with special reference to the origin of the genus *Oryza*. In: Rice is life: scientific perspectives for the 21st century (eds. K. Toriyama, K.L. Heong & B. Hardy). *Proceedings of the World Rice Research Conferenc*e, Tsukuba, Japan, pp. 40 – 44.
4. Haider S., Ballester B., Smedley D., Zhang J., Rice P. & Kasprzyk A. (2009). BioMart Central Portal-unified access to biological data. *Nucleic Acids Research* **37**: 23 – 27. DOI: https://doi.org/10.1093/nar/gkp265
5. Hillis D.M. & Bull J.J. (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* **42**(2): 182 – 192. DOI: https://doi.org/10.1093/sysbio/42.2.182
6. Itoh H. & Izawa T. (2013). The coincidence of critical day length recognition for florigen gene expression and floral transition under long-day conditions in rice. *Molecular Plant* **6**(3): 635 – 649. DOI: https://doi.org/10.1093/mp/sst022
7. Jukes T.H. & Cantor C.R. (1969). Evolution of protein molecules. *Mammalian Protein Metabolism* **3**(21): 132. DOI: https://doi.org/10.1016/b978-1-4832-3211-9.50009-7
8. Khush G.S. (1997). Origin, dispersal, cultivation and variation of rice. *Plant Molecular Biology* **35**(1-2): 25 – 34. DOI: https://doi.org/10.1023/A:1005810616885
9. Khush G.S. & Brar D.S. (2001). Rice genetics from Mendel to functional genomics. In: Rice genetics IV (eds. G.S. Khush, D.S. Brar & B. Hardy), *Proceedings of the Fourth International Rice Genetics Symposium*, Los Banos,

The Philippines, Science Publishers Inc., New Hampshire, USA, International Rice Research Institute, Los Banos, The Philippines, pp. 3 – 25.
10. Komiya R., Ikegami A., Tamaki S., Yokoi S. & Shimamoto K. (2008). *Hd3a* and *RFT1* are essential for flowering in rice. *Development (Cambridge, England)* **135**: 767 – 774. DOI: https://doi.org/10.1242/dev.008631
11. Koonin E.V. (2005). Orthologs, paralogs, and evolutionary genomics. *Annual Review of Genetics* **39**: 309 – 338. DOI: https://doi.org/10.1146/annurev.genet.39.073003.114725
12. Librado P. & Rozas J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**(11): 1451 – 1452. DOI: https://doi.org/10.1093/bioinformatics/btp187
13. Liu W. *et al*. (12 authors) (2015). IBS: an illustrator for the presentation and visualization of biological sequences. *Bioinformatics* **31**(20): 3359 – 3361. DOI: https://doi.org/10.1093/bioinformatics/btv362
14. Long M., Rosenberg C. & Gilbert W. (1995). Intron phase correlations and the evolution of the intron/exon structure of genes. *Proceedings of the National Academy of Sciences* **92**(26): 12495 – 12499. DOI: https://doi.org/10.1073/pnas.92.26.12495
15. Lu L., Yan W., Xue W., Shao D. & Xing Y. (2012). Evolution and association analysis of *Ghd7* in rice. *PloS ONE* **7**(5): 34021. DOI: https://doi.org/10.1371/journal.pone.0034021
16. Mohanty S. (2013). Trends in global rice consumption. *Rice Today*. Available at: *http://irri.org/rice-today/trends-in-global-rice-consumption*, Accessed 21 August 2016.
17. Monaco M.K. *et al*. (26 authors) (2014). Gramene 2013: comparative plant genomics resources. *Nucleic Acids Research* **42**: D1. DOI: https://doi.org/10.1093/nar/gkt1110
18. Nei M. (1987). *Molecular Evolutionary Genetics*, p. 512. Columbia University Press, New York, USA.
19. Pamilo P. & Nei M. (1988). Relationships between gene trees and species trees. *Molecular Biology and Evolution* **5**(5): 568 – 583.
20. Rakshit S., Rakshit A., Matsumura H., Takahashi Y., Hasegawa Y., Ito A., Ishii T., Miyashita N.T. & Terauchi R. (2007). Large-scale DNA polymorphism study of *Oryza sativa* and O. *rufipogon* reveals the origin and divergence of Asian rice. *Theoretical and Applied Genetics* **114**(4): 731 – 743. DOI: https://doi.org/10.1007/s00122-006-0473-1
21. Shrestha R., Gómez-Ariza J., Brambilla V. & Fornara F. (2014). Molecular control of seasonal flowering in rice, *Arabidopsis* and temperate cereals. *Annals of Botany* **114**(7): 1445 – 1458. DOI: https://doi.org/10.1093/aob/mcu032
22. Stamatakis A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**(21): 2688 – 2690. DOI: https://doi.org/10.1093/bioinformatics/btl446
23. Tan J. *et al.* (16 authors) (2016). *OsCOL10*, a CONSTANS-like gene, functions as a flowering time repressor downstream of *Ghd7* in rice. *Plant and Cell Physiology* **57**(4): 798 – 812. DOI: https://doi.org/10.1093/pcp/pcw025
24. The UniProt Consortium (128 collaborators) (2014). UniProt: a hub for protein information. *Nucleic Acids Research* **43**: D204 – D212.
25. Thompson J.D., Higgins D.G. & Gibson T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**(22): 4673 – 4680. DOI: https://doi.org/10.1093/nar/22.22.4673
26. Vaughan D.A., Morishima H. & Kadowaki K. (2003). Diversity in the *Oryza* genus. *Current Opinion in Plant Biology* **6**(2): 139 – 146.
27. Wang J., He X., He Z., Wang H. & Xia X. (2009). Cloning and phylogenetic analysis of *phytoene synthase 1* (*Psy1*) genes in common wheat and related species. *Hereditas* **146**(5): 208 – 256. DOI: https://doi.org/10.1111/j.1601-5223.2009.02132.x
28. Wang M. *et al*. (31 authors) (2014). The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nature Genetics* **46**(9): 982 – 988. DOI: https://doi.org/10.1038/ng.3044
29. Weng X., Wang L., Wang J., Hu Y., Du H., Xu C., Xing Y., Li X., Xiao J. & Zhang Q. (2014). *Grain number, plant height, and heading date7* is a central regulator of growth, development, and stress response. *Plant Physiology* **164**(2): 735 – 747. DOI: https://doi.org/10.1104/pp.113.231308
30. Xue W. *et al*. (11 authors) (2008). Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. *Nature Genetics* **40**(6): 761 – 767. DOI: https://doi.org/10.1038/ng.143
31. Yang L., Liu T., Li B., Sui Y., Chen J., Shi J., Wing R.A. & Chen M. (2012). Comparative sequence analysis of the *Ghd7* orthologous regions revealed movement of *Ghd7* in the grass genomes. *PloS ONE* **7**(11): 50236. DOI: https://doi.org/10.1371/journal.pone.0050236