Ibro, J Rice Res 2022, 10:7

Rice Research: Open Access

ISSN: 2375-4338

Research Article

Open Access

A Research on Rice genomes Recorded Ancient Pararetrovirus Activities

Germaine Ibro*

Department of Soil Science, School of Agricultural Sciences, University of Zambia, Zambia

Abstract

Viral fossils in rice genomes are a best entity to understand ancient pararetrovirus activities through host plant history because of our advanced knowledge of the genomes and evolutionary history with rice and its related species. Here, we explored organization, geographic origins and genealogy of rice pararetroviruses, which were turned into endogenous rice tungro bacilliform virus-like (eRTBVL) sequences. About 300 eRTBVL sequences from three representative rice genomes were clearly classified into six families. Most of the endogenization events of the eRTBVLs were initiated before differentiation of the rice progenitor (> 160,000 years ago). We successfully followed the genealogy of old relic viruses during rice speciation, and inferred the geographical origins.

Keywords: Paleovirology; Endogenous pararetrovirus; Rice; Genealogy; Recombination

Introduction

Pararetroviruses (*Caulimoviridae* and *Hepadnaviridae*) have a double-stranded DNA genome and resemble retroviruses in that they employ reverse transcription for genome replication, but lack the processes and molecular machinery for integration into the host genome (Temin, 1985). However, with the increasing decryption of plant genomes, a growing number of endogenous pararetrovirus (EPRV) sequences have been discovered as ancient integrated analogs of most members of the *Caulimoviridae* family,. This incidental integration of pararetroviruses was thought to involve illegitimate recombination or double-strand break repair via non-homologous end joining between the pararetrovirus and host genomes.Rice tungro disease is a significant constraint on rice (*Oryza sativa* L.) yields in South and South-East Asia. The disease is caused primarily by infection with rice tungro bacilliform virus (*Caulimoviridae, Tungrovirus*; RTBV), the known extant rice pararetrovirusThe evolutionary history of rice has been studied deeply and described. Cultivated rice, which includes two subspecies *indica* and *japonica*, was domesticated from the wild rice species *O. rufipogon* . *O. rufipogon* is thought to have diverged into at least two ecotypes, perennial and annual (the annual type is also called *O. nivara*), 0.16 million years ago (Mya), as estimated by the molecular clock.Here, we compared the eRTBVL sequences from the three rice genome databases by comprehensive *in silico* analyses and screened orthologous sequences in diverse cultivated and wild rice accessions, to understand the evolutionary route of the eRTBVL sequences.

To investigate the origin of the endogenous pararetrovirus elements in the rice genome, we performed a strict *in silico* BLAST search (e value <1e−10) of the eRTBVL sequences against a recently released draft genome of the specific *O. rufipogon* line (W1943) (Huang et al., 2012) that is distributed in eastern China (Lu et al., 2008). Among the 1607 contigs that were found to be harboring eRTBVL segments, the 206 contigs with>50 bp that flanked at least one end the eRTBVL sequences (comprising more than 103 eRTBVL loci in this *O. rufipogon* genome) were extracted and mapped to the *japonica* (Nipponbare) and *indica* (93-11) genomes for further examination. After mapping, 57 of the eRTBVL elements in W1943 were anchored successfully with unique and unambiguous matches to the Nipponbare and 93-11 genomes, including the elements that were shared with Nipponbare and/or 93-11 (species-shared eRTBVLs), and W1943-specific eRTBVLs.

Materials and Method

In contrast, five 93-11-specific elements were observed in a wide range of *indica* accessions and the *O. rufipogon* accessions from South East, South Asia, and China, but were absent in the *japonica* accessions . Four W1943-specific elements were absent in all other accessions. Hence, the eRTBVL pool for *japonica* might have originated from a Chinese *O. rufipogon* population (but not from the line of W1943), and the eRTBVL pool for *indica* could be mixed with those of *O. rufipogon* populations from South East and South Asia in addition to China. Interestingly, the Nipponbare/93-11-shared elements (except the eRTBVL-D elements) could have been introduced into both *japonica* and *indica* cultivars only through Chinese *O. rufipogon* (perennial) lineages, and not through *O. rufipogon* (annual) lineages in South East or South Asia, which did not possess these Nipponbare/93-11-shared eRTBVL elements. The eRTBVLs (except eRTBVL-D) from *O. rufipogon* (annual) lineages in South East and South Asia were distributed mostly toward *indica* cultivarindicating that eRTBVL-D could be the oldest family among them. Because of the apparent clustering that was observed within the eRTBVL families, the four distinct families must have.

Data Analysis

eRTBVL families identified by a phylogenetic analysis based on the RT/RH region in eRTBVL sequences. The maximum likelihood (ML) phylogenetic tree was constructed using the nucleotide sequences of the RT/RH region of eRTBVLs from the *O. rufipogon* (W1943), *japonica* (Nipponbare), and *indica* (93-11) genomes (Only sequences with lengths>80% of the length of RT/RH consensus sequence were used; therefore, the orthologous sequences of JaE11-2 that we had collected were not included because they contained deletions.). The genomic locations of the eRTBVL sequences are shown in each taxon named as ID, chromosome/contig, position, and strand. The rice genomes to which the sequences belong are labeled with colored solid circles in front of the taxon names (orange, Nipponbare; green, 93-11;

***Corresponding author:** Germaine Ibro, Department of Soil Science, School of Agricultural Sciences, University of Zambia, Zambia; E-mail: germaine.ibro27@gmail.com

Received: 01-Jul-2022, Manuscript No. rroa-22-70997; **Editor assigned:** 04-Jul-2022, PreQC No. rroa-22-70997 (PQ); **Reviewed:** 18-Jul-2022, QC No. rroa-22-70997; **Revised:** 22-Jul-2022, Manuscript No. rroa-22-70997(R); **Published:** 29-Jul-2022, DOI: 10.4172/2375-4338.1000311

Citation: Ibro G (2022) A Research on Rice genomes Recorded Ancient Pararetrovirus Activities. J Rice Res 10: 311.

Copyright: © 2022 Ibro G. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

blue, W1943). Sequences sorted into the same family are indicated by the same background color. The eRTBVL-X sequences are indicate. Each of the *japonica*, *indica*, and *O. rufipogon* genomes harbored about 100 copies of the eRTBVL-A, -B, -C, and -D families (this study, and Liu et al., 2012), but in different proportions. We examined the endogenization patterns of the eRTBVLs in each of the three genomes by comparing their relative abundances in the eRTBVL-A, -B, and -C families. To maximize the number of eRTBVL segments and at the same time exclude the short ones with ambiguous taxonomy, all the sequences with lengths>60% of the lengths of the RT/RH region, ORFz and IGR (the ML trees based on these regions showed clear distant clusters of the eRTBVL-A, -B and -C families) were retrieved from the three rice genomes for the endogenization pattern analysis. To estimate the relative abundance of the eRTBVL-A, -B and -C families in the three genomes, the segments that belonged to the eRTBVL-D and -X families or had ambiguous taxonomy were not taken into account. In the same way, we also compared the relative abundance of the species-specific eRTBVLs in the three families after removing the known species-shared elements.

Results and Discussion

To examine the relationship between rice ecotypes and the eRTBVL families, orthologous sequences of the eRTBVL loci that were identified in the three rice genome databases were further explored in a core collection of 65 cultivated (16 *japonica* and 49 *indica*) and 30 wild. We screened orthologous eRTBVL insertions by PCR with plural primer pairs that were designed to amplify the eRTBVL sequence interior to the flanking sequences. Because of the abundant repetitive sequences in the flanking regions of a number of eRTBVL loci, the orthologs of only 34 eRTBVL loci, representing Nipponbare/93-11-shared and -specific loci and W1943-specific loci, were amplified six of the eRTBVL-D elements that we examined had been fixed in almost all the cultivated and wild accessions in the collection, which confirmed eRTBVL-D as the most ancient family. Seven Nipponbare/93-11-shared eRTBVL-A, -B, and -C elements, including unsigned elements, were distributed widely in the cultivated rice accessions and in the Chinese *O.* Nipponbare/93-11-shared elements (except the eRTBVL-D elements) could have been introduced into both *japonica* and *indica* cultivars only through Chinese *O. rufipogon* (perennial) lineages, and not through *O. rufipogon* (annual) lineages in South East or South Asia, which did not possess these Nipponbare/93-11-shared eRTBVL elements. The eRTBVLs (except eRTBVL-D) from *O* (Table 1).

In contrast, five 93-11-specific elements were observed in a wide range of *indica* accessions and the *O. rufipogon* accessions from South East, South Asia, and China, but were absent in the *japonica* accessions . Four W1943-specific elements were absent in all other accessions. Hence, the eRTBVL pool for *japonica* might have originated from a Chinese *O. rufipogon* population (but not from the line of W1943), and the eRTBVL pool for *indica* could be mixed with those of *O. rufipogon* populations from South East and South Asia in addition to China. Interestingly, the *rufipogon* (annual) lineages in South East and South Asia were distributed mostly toward *indica* cultivars The eRTBVL-D family was identified as the oldest family distinct from the younger eRTBVL-A, -B and -C families . Each of the eRTBVL-D sequences had orthologs in nearly all the *O. sativa* and *O. rufipogon* accessions examined here . The viral eRTBVL-D sequences, therefore, had become immobilized in the genomes of the progenitor of *O. rufipogon*, at least since 0.16 Mya when *O. rufipogon* began to diverge (Zheng and Ge, 2010). The other viral lineages of the eRTBVL families emerged relatively recently, after the divergence of the *O. rufipogon* ecotypes . The nucleotide divergences after endogenization suggested that the emergence of the viral lineage of eRTBVL-B sequences could have preceded those of eRTBVL-A, -C and -X sequences Although an alternative scenario (the virus lineage of eRTBVL-B was younger than the lineages of eRTBVL-A and -C, and was a recombinant from the latters) cannot be completely excluded, the current. (Figure 1). B is reasonable and congruent with our observations *Japonica* and *indica* were thought to be domesticated in different processes; *japonica* was established mainly from perennial *O. rufipogon* in southern China, while *indica* was a subsequent hybrid of the ancient *japonica* and the annual *O. rufipogon* in South East and South Asia. The genomic data of Nipponbare, 93-11, and W1943 were downloaded from the RAP build 5 (Rice Annotation BGI-RIS V2 and RiceHap3 databases, respectively. To identify eRTBVLs in W1943, a BLASTN search was performed using the BLAST+ 2.2.27 application with the were kept for mapping. Mapping of the W1943 eRTBVLs onto the *japonica* and *indica* genomes by BLASTN was based on a maximum of 10 kb of flanking sequence(s) of each eRTBVL locus in the filtered contigs. The unambiguously mapped eRTBVLs were compared with their counterparts in the *japonica* and *indica* genomes and those that mapped to unique genomic retrieved. The phylogenetic trees were constructed based on the defined regions in the eRTBVL sequences. All the segments with lengths>80% of the lengths of each of the regions were retrieved from the Nipponbare, 93-11 and W1943 genomes. The sequences were aligned using followed by manual editing, and ambiguous regions and missing data were removed. Maximum likelihood (ML) phylogenetic analysis was performed by MEGA5. el for the MP, CP and PR regions; T92+G model for the RT/RH region, ORFz and IGR; T92 model for ORFx). ML analysis of

Table 1: Classification of eRTBVL family.

DATASET	eRTBVL family	RT/RH region		ORFz		IGR	
		Divergenceb	Sample sizee	Divergenceb	Sample size	Divergenceb	Sample size
Nipponbare	A	0.0024	14	0.0028	14	0.0056/0.0044	09-Jul
	B	0.0055	10	0.0084	8	0.0102	13
	C	0.0033	11	0.0019	11	0.0049	10
93-11	A (A1/A2)d	0.005	4	0.0053	4	NA / NA	NA/NA
	B	0.0057	13	0.0087	11	0.0089	13
	C	0.0043	6	0.0037	7	0.0031	7
W1943	A (A1/A2)d	NA	NA	0.0047	7	0.0113/NA	4/NA
	B	0.0062	10	0.0095	12	0.0122	9
	C	0.006	6	0.0027	10	0.0056	16
Combined	A (A1/A2)d	0.0014	21	0.0011	25	0.0042/0.0039	14-Oct
	B	0.003	33	0.0051	31	0.0051	35
	C	0.0023	23	0.0015	28	0.0039	33

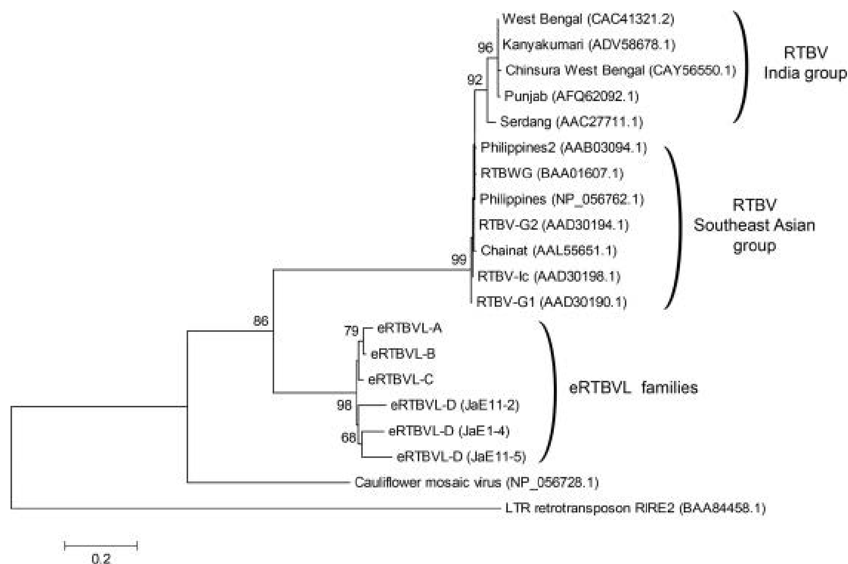

Figure 1: Distribution of eRTBAL in Aisa.

the eRTBVL families and RTBV strains was based on the amino acid sequences of the RT/RH region using the WAG+G model, which was chosen after model testing. Support for the ML trees was evaluated by 1000 bootstrap replicates. All the sequence alignments

Conclusion

Total DNAs were extracted from the leaf samples of the cultivated and wild rice accessions by Plant DNAzol (Invitrogen, Carlsbad, CA, USA). The DNA concentration in each sample was measured by NanoDrop 2000 (Thermo Fisher Scientific, Wilmington, DE, USA) and all were adjusted to a similar level. Primer pairs were generally designed and used as follows: for amplification 1, a forward primer for the left flank (P1) and a reverse primer for the interior of the eRTBVL locus (P2) were used; for amplification 2, a forward primer for the interior of the eRTBVL locus (P3) and a reverse primer for the right flank (P4) were used; for amplification 3, a primer pair for left and right flanks (P1, P4) were used to amplify the whole length of the eRTBVL locus or the empty donor site. Polymerase chain reactions (PCR) were performed using Ex Taq or LA Taq polymerase (Takara, Shiga, Japan) in a PTC-200 thermal cycling system (GMI, Ramsey, MN, USA). When the whole length of some long eRTBVLs failed to amplify, the reaction was adjusted to amplify the possible empty donor sites in the corresponding rice accessions. The PCR products were resolved on a 1–2% agarose gel and analyzed with a Typhoon 8600 PhosphorImager (GE Healthcare, Little Chalfont, U.K.).Consensus sequences of the eRTBVL-A1, -A2, -B, -C and -X families were acquired by aligning the nucleotide sequences from the defined regions for each family (the chimeric sequences that possibly resulted from intra-recombination after integration were excluded). The eRTBVL sequences that spanned at least 80% of the length of each region in the three rice genomes were used in the alignments. The consensus sequences of each region were combined, and ambiguous regions at the terminals were trimmed to produce the final whole length consensus sequences. These consensus sequences were then aligned by ClustalW and manually edited for use in the recombination analysis. We employed the RDP, GENECONV and BOOTSCAN methods in the RDP3 package. eRTBVL sequences from the Nipponbare genome DDBJ accession numbers: BR000029-BR000031as queries. The following BLAST parameters were used: word size, 11; gap open, 5; gap extend, 2; penalty, −3; and reward, 2. The contigs spanning high-identity matches (e-values<1e−10, alignment length>100 bp) were extracted, and filtered so that only the contigs with>50 bp flanking at least one end of the eRTBVL loci.

References

1. Zhou Z, Robards K (2002) Composition and functional properties of rice. Int J Food Sci Technol 37: 849-868.
2. Yoshida H, Tomiyama Y, Mizushina Y(2010) Lipid components, fatty acids and triacylglycerol molecular species of black and red rices. Food Chem 123: 210-215.
3. Kim NH, Kwak J, Baik JY, Yoon MR, Lee JS, et al. (2015) Changes in lipid substances in rice during grain development. Phytochemistry 116: 170-179.
4. Heinemann RJB, Fagundes PDL (2005) Comparative study of nutrient composition of commercial brown, parboiled and milled rice from Brazil. J Food Compo Anal 18: 287-296.
5. Chen HH, Hung CL (2015) Effect of low-pressure plasma exposure on the storage characteristics of brown rice. Food Bioprocess Technol 8: 471–477.
6. Stout MJ, Riggio MR, Zou L, Roberts R(2002) Flooding influences ovipositional and feeding behaviour of the rice water weevil, Lissorhoptrus oryzophilus (Coleoptera: Curculionidae). J Econ Entomol 95: 715-721.
7. Zou L, Stou MJ (2004) Degree-day models to predict the emergence and development of rice water weevil (Coleoptera: Curculionidae) in southwest Louisiana. Environ Entomol 33: 1541-1548.
8. Zou L, Stout MJ (2004) The effects of feeding by the rice water weevil, Lissorhoptrus oryzophilus Kuschel on the growth and yield components of rice, Oryza Sativa. Agric For Entomol 6: 47-53.
9. Stout MJ, Hamm JC(2011) Comparison of the efficacies of cruiser maxx and dermacorX-100 against the rice water weevil at three planting dates. Annu Res Report 103: 331-332.
10. Blackman B, Autin T (2014) Management practices of Louisiana and Texas rice growers. La Agric Mag 57: 14-15.