



Continuous infiltration and evolutionary trajectory of nuclear organelle DNA in *Oryza*

Chenbo Gong, Yicheng Huang, Mengmeng Liu, et al.

Genome Res. 2025 35: 1349-1363 originally published online April 25, 2025

Access the most recent version at doi:[10.1101/gr.279609.124](https://doi.org/10.1101/gr.279609.124)

References This article cites 96 articles, 13 of which can be accessed free at:
<http://genome.cshlp.org/content/35/6/1349.full.html#ref-list-1>

Creative Commons License This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <https://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

CRISPR and RNAi Genetic Screening.
Your new superpower.

LEARN MORE



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

© 2025 Gong et al.; Published by Cold Spring Harbor Laboratory Press

Research

Continuous infiltration and evolutionary trajectory of nuclear organelle DNA in *Oryza*

Chenbo Gong,¹ Yicheng Huang,¹ Mengmeng Liu,¹ Yong Zhou,² Yinwei Xu,¹ Nahed Mohammed,² Xintong Qiao,¹ Andrea Zuccolo,² Weibo Xie,¹ Rod A. Wing,^{2,3} Jianwei Zhang,¹ Fei Zhou,¹ and Yongjun Lin¹

¹National Key Laboratory of Crop Genetic Improvement, Hubei Hongshan Laboratory, Huazhong Agricultural University, Wuhan 430070, China; ²Center for Desert Agriculture, Biological and Environmental Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia; ³Arizona Genomics Institute, School of Plant Sciences, University of Arizona, Tucson, Arizona 85721, USA

Transfer of chloroplast or mitochondrial DNA into the nuclear genome is a common phenomenon in many species. However, little is known about the evolutionary fate and mechanism of transfer of organellar DNA sequences in higher plants. We observe abundant insertions of organelle DNA into the nuclear genomes of 22 genome assemblies across seven *Oryza* species and further categorize nuclear organelle DNA (NORG) into 3406 orthologous groups. Analysis of the whole-genome resequencing data from 3458 *O. sativa*, *O. glaberrima*, and *O. barthii* accessions indicate that NORGs have intra- and inter-population variability owing to sequence loss and transposon insertion during evolution. Our results also suggest that NORGs have been continuously produced during the evolution of *Oryza*, and both double-strand break repair pathways and replication-based mechanisms play important roles in integrating organelle DNA into the nuclear genome. Further investigation indicates that complex NORGs are formed through single mutational events before or during the insertion process via ligation of multiple plastid and/or mitochondrial DNA with each other. In summary, this work provides novel insights into the process of endosymbiotic DNA transfer and its role in reshaping genome variation and plant genome evolution.

[Supplemental material is available for this article.]

The observation of “promiscuous DNA,” which is defined as a nucleotide sequence occurring in more than one of the three membrane-bound organellar genetic systems of eukaryotic cells (Ellis 1982), provides evidence for DNA transfer between organelles and the nucleus. Such transfers of mitochondrial DNA (mtDNA) or plastid DNA (ptDNA) to the nucleus result in formation of nuclear mitochondrial DNA (NUMT) and nuclear plastid DNA (NUPT), which are collectively known as nuclear organelle DNA (NORG) (Leister 2005). The consistent and frequent occurrence of organelle-to-nucleus transfer of DNA has been demonstrated by many experimental studies (Thorsness and Fox 1990; Huang et al. 2003; Stegemann et al. 2003).

To date, NUMT has been ubiquitously found in eukaryotic nuclear genomes (Richly 2004; Hazkani-Covo et al. 2010; Dayama et al. 2020), representing an important source of genetic variation in the nuclear genome, and plays an important role in cellular evolution (Liang et al. 2018; Puertas and González-Sánchez 2020). In humans, de novo formation of NUMT occurs approximately once in every 4000 births (Wei et al. 2022), with a higher rate in cancer cells (Ju et al. 2015; Wei et al. 2022), leading to high NUMT diversity across the human population. The insertion of mtDNA into the nuclear genome can cause genomic instability via mutations or disruption of normal gene function, which may result in several human diseases such as carcinogenesis

and genetic disorders (Puertas and González-Sánchez 2020; Harutyunyan 2024).

Transfer of mtDNA and ptDNA to the nucleus occurs frequently in plants (Huang et al. 2004; Richly and Leister 2004; Noutsos et al. 2005; Portugez et al. 2018; Zhang et al. 2020). Moreover, the large size, complex structure, and frequent rearrangement events of mtDNA (Gualberto and Newton 2017; Morley and Nielsen 2017; Wu et al. 2022) contribute to the diversity of NORGs in plants. Plant NORG diversity is further increased by the mosaic sequences formed from rearranged DNA originating from both mtDNA and ptDNA (Noutsos et al. 2005; Portugez et al. 2018). Triticeae species tend to have more nuclear organellar genes than other Poaceae species (Chen et al. 2023). In Asian cultivated rice, several NUPT and NUMT events occur specifically in one subspecies (Huang et al. 2005; Guo et al. 2008; Wang and Timmis 2013), and giant NORGs (>10 kb) vary greatly in Asian and African rice (Ma et al. 2020). Currently, little is known about the overall profile and dynamics of NORGs in different *Oryza* species or across diverse populations.

Several potential mechanisms for the formation of NORGs have been reported. Capture of mtDNA and ptDNA during double-strand break repair (DSBR) has been confirmed in yeast (Ricchetti et al. 1999; Yu and Gabriel 1999; Decottignies 2005) and tobacco (Wang et al. 2012, 2018a). In primates, NUMT exhibits the characteristics of non-homologous end joining (NHEJ)

Corresponding authors: rod.wing@kaust.edu.sa, jzhang@mail.hzau.edu.cn, zhoufei@mail.hzau.edu.cn, yongjunlin@mail.hzau.edu.cn

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.279609.124>.

© 2025 Gong et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <https://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

(Hazkani-Covo and Covo 2008; Wei et al. 2022). However, nuclear integration of mtDNA in somatic cells frequently occurs together with other complex rearrangement events, suggesting that formation of NUMTs may involve mechanisms in addition to NHEJ (Ju et al. 2015). In plants, the specific mechanism underlying NORG formation remains elusive.

The *Oryza* genus is a grass model system for studying molecular evolution. With an evolutionary history of ~15 million years, *Oryza* includes two cultivated rice species, Asian cultivated rice *Oryza sativa* and African cultivated rice *Oryza glaberrima*, along with 25 wild species. It comprises six diploid genome types (AA, BB, CC, EE, FF, and GG) and five allopolyploid genome types (BBCC, CCDD, HHJJ, HHKK, and KKLL) (Wing et al. 2018). Here, to obtain the landscape of NORGs and understand its variations during the evolution of *Oryza*, we identified and analyzed NORGs at the pangenome level focusing on evolutionary patterns, population dynamics, formation, and fragmentation mechanisms of NORGs. The results provide a resource for interpreting NORG variation and formation across diverse rice populations and improve our understanding of genome evolution.

Results

Abundant organelle DNA insertions in the nuclear genomes of *Oryza* species

To obtain the landscape of NORGs and understand its continuous variation during the evolution of *Oryza* genus, we identified and analyzed NORGs at the pangenome level using 22 high-quality genomes of representative Asian cultivated rice (*O. sativa*) (Kawahara et al. 2013; Zhou et al. 2020; Song et al. 2021) and six other *Oryza* species (Table 1; Supplemental Fig. S1; Stein et al. 2018; Ntakitimana et al. 2023; Thathapalli Prakash et al. 2023). In total, we identified 21,971 NORGs in all 22 genomes. Among these NORGs, 20,085 NORGs in the AA genome were classified into 3406 orthologous NORG groups (ONGs), whereas only 948 ONGs were identified in the IRGSP-1.0 *O. sativa* ssp. *japonica* (AA genome) reference genome (Supplemental Table S1). This discrepancy indicates that simply relying on a single reference genome fails to comprehensively represent the diversity of NORGs within the *Oryza* genus. A total of 138 to 841 kb NUPTs and 180 to 932 kb NUMTs were identified (Table 1; Supplemental Table S2). The total lengths of NORGs across the 22 genomes range from 301 kb to 1574 kb, accounting for 0.12% to 0.42% of the nuclear genomes, with an average of 0.26% in the AA genome (Table 1; Supplemental Fig. S1). All NUPTs in each genome cover the complete plastid genome region, indicating that the ptDNA sequences are ubiquitously transferred to and retained in the nuclear genome (Supplemental Fig. S2), whereas NUMTs do not cover the entire mitochondrial genome (covering 43% to 79% of the mtDNA) (Supplemental Fig. S3). The origins of NUPTs and NUMTs from organellar genomes were not random, with the *RPS8* gene region of ptDNA and the *RRN5–RRN18* gene region of mtDNA appearing the most frequently (Supplemental Figs. S2, S3).

The number of ONGs in each genome is similar among different accessions and has a positive correlation with the chromosome length ($R^2 = 0.759$) (Fig. 1A,B), but the total length of NORGs demonstrates a relatively weaker association with chromosome length ($R^2 = 0.0303$) (Fig. 1C,D). The majority of ONGs (83%) are <1000 bp, indicating that NORG insertion usually involves short fragments (Fig. 1E; Supplemental Fig. S4). ONGs appear to be uniformly distributed across chromosomes with no definite tendency for

the insertion into centromeric regions (Kolmogorov–Smirnov test P -value > 0.01 in 10,000 uniformly distributed simulations) (Fig. 1F; Supplemental Fig. S5). Still, ONGs tend to be located in noncoding regions of the nuclear genome, with only a small number of ONGs overlapping with exons in each genome (chi-square test P -value < 0.01 in 1000 NORG random shuffle simulations; see Methods) (Supplemental Table S3). In plants, some species accumulate longer NUMT than NUPT with genome size as a key factor determining accumulated length of relatively older NUPT and NUMT (Yoshida et al. 2017). The presence of giant NORGs (>10 kb), such as those in Chr 10 and Chr 12 of Nipponbare and CHAO MEO, respectively, might be the reason for the variation in the total length of NORGs.

In total, we classified five types of NORGs according to the origin of organelle DNA (Fig. 2A–E). The proportions of different types of NORGs were similar in all genomes (Fig. 2F). Among the NORGs with lengths <500 bp, simple NUMTs and simple NUPTs predominated. In contrast, for NORGs >1000 bp in length, the majority of them were complex NORGs, with sequences matching multiple distinct regions from plastid, mitochondria, or both organelles (Fig. 2G). The three types of complex NORGs also had different numbers of fragments, with nuclear mosaic insertions (NUMINs) (Fig. 2E) having more fragments (Fig. 2H).

Analysis of near-intact organellar genes (NIOGs; $\geq 95\%$ coverage of an organellar gene) within the nuclear genome revealed that 435 (13%) ONGs contain NIOGs (Supplemental Table S6). The NIOG number ranges from 308 to 969, covering 123 (58%) to 189 (89%) organellar genes (Supplemental Table S7). Among these, 34% to 59% of the protein-coding genes lack complete CDS (Supplemental Table S6). The number of nuclear mitochondrial genes was smaller than that of nuclear plastid genes, probably because mitochondrial genomes have a lower gene density than plastids (Supplemental Table S7).

Recent transfer and nonconservation of NORGs

Abundant presence/absence variations (PAVs) of ONGs were identified in 20 AA genome accessions (Fig. 3A). Among the 1728 ONGs in *O. sativa* genomes, 34% (590/1728) of them are present in all the 16 *O. sativa* accessions, suggesting that these ONGs that were already present in the common ancestor of *O. sativa* are termed the *O. sativa* “common ONGs” (Fig. 3B). Among these “common *O. sativa* ONGs,” ~80% (470) of them are shared with both *O. glaberrima* and *Oryza barthii* (Fig. 3C). The remaining 66% of ONGs in *O. sativa* have PAVs (Fig. 3B). Abundant species-specific ONGs were found, suggesting that the insertion of organelle DNA into nuclear genomes is ongoing (Fig. 3D).

Interestingly, the more ancient the evolutionary branch, the fewer shared ONGs were identified. Five hundred ninety ONGs were conserved in all 16 *O. sativa* accessions; 180 ONGs were present in all 20 accessions of the AA genome, of which only 28 ONGs were present in *Oryza punctata* (BB genome) (Fig. 3D). This indicates that ONGs were not conserved during evolution, and most of ONGs identified in this study occurred as new insertions after the speciation between ancestors of AA and BB genome species.

To understand the characteristics of conserved ONGs, the ONGs in Nipponbare were divided into three groups, including those fixed in AA species, those fixed in *O. sativa*, and those not fixed in *O. sativa*. The distribution of ONG positions relative to genes was then compared. Conserved ONGs are more probably located within genic regions than nonconserved ONGs (Fisher's exact test P -value = 5×10^{-4}) (Supplemental Fig. S6), indicating a

Table 1. Summary of NORG content in assemblies of *Oryza* genus

Species and variety	Subpopulation ^a	Genome size (Mb)	NORG length (kb)	Proportion to nuclear genome	NUPT length (kb) ^b	NUMT length (kb) ^b	Nuclear genome info	Genome type
<i>O. sativa</i> (Asian cultivated rice)								
N22	cA	382.954	1011.926	0.26%	536 (51%)	510 (49%)	OsN22RS2	AA
NATEL BORO	cA2	383.721	868.398	0.23%	474 (55%)	394 (45%)	Os127652RS1	AA
ARC 10497	cB	378.464	818.082	0.22%	523 (64%)	296 (36%)	Os117425RS1	AA
CHAO MEO	GJ-subtrp	376.857	1574.855	0.42%	780 (50%)	796 (50%)	Os132278RS1	AA
Nipponbare	GJ-temp	373.246	1401.968	0.38%	836 (58%)	598 (42%)	IRGSP-1.0	AA
Azucena	GJ-trop1	379.628	1105.439	0.29%	728 (66%)	378 (34%)	AzucenaRS1	AA
KETAN NANGKA	GJ-trop2	380.759	864.677	0.23%	529 (61%)	337 (39%)	Os128077RS1	AA
Zhenshan97	XI-1A	387.326	920.786	0.24%	496 (52%)	460 (48%)	ZS97RS3	AA
IR 64	XI-1B1	386.699	763.879	0.20%	391 (51%)	374 (49%)	OsiR64RS1	AA
PR 106	XI-1B2	391.176	951.257	0.24%	419 (44%)	537 (56%)	Os127742RS1	AA
GOBOL SAIL	XI-2A	391.773	855.828	0.22%	360 (42%)	494 (58%)	Os132424RS1	AA
LARHA MUGAD	XI-2B	390.196	828.319	0.21%	377 (46%)	450 (54%)	Os125619RS1	AA
LIMA	XI-3A	392.625	932.976	0.24%	366 (39%)	566 (61%)	Os127564RS1	AA
KHAO YAI GUANG	XI-3B1	393.738	938.22	0.24%	351 (38%)	586 (62%)	Os127518RS1	AA
LIU XU	XI-3B2	392.033	932.67	0.24%	475 (51%)	458 (49%)	Os125827RS1	AA
Minghui63	XI-adm	387.424	845.852	0.22%	372 (41%)	530 (59%)	MH63RS3	AA
<i>O. glaberrima</i> (African cultivated rice)		344.458	948.328	0.28%	330 (33%)	660 (67%)	OglaRS2	AA
<i>O. barthii</i>		344.915	1200.189	0.35%	313 (25%)	932 (75%)	ObarRS3	AA
<i>O. glumaepatula</i>		386.181	1144.465	0.30%	841 (71%)	336 (29%)	OgluRS3	AA
<i>O. meridionalis</i>		379.256	1171.963	0.31%	561 (46%)	657 (54%)	OmerRS3	AA
<i>O. punctata</i>		419.078	698.815	0.17%	297 (40%)	449 (60%)	OpunRS2	BB
<i>O. brachyantha</i>		258.904	301.256	0.12%	138 (43%)	180 (57%)	ObraRS2	FF

^aSubpopulations are as follows: (cA) *circum-Aus*; (cB) *circum-Basmati*; (GJ) *Geng-japonica*, for which (trop) tropical, (subtrp) subtropical, and (temp) temperate; and (XI) *Xian-indica*.

^bBecause of the overlap of some NUMTs and NUPTs and interference from mitochondrial ptDNA, the total length of NUPTs and NUMTs may exceed that of NORGs.

potential evolutionary constraint on the retention of ONGs during evolution.

Population dynamics of NORG insertions in Asian and African rice

To further investigate the distribution of ONGs in Asian and African rice at the population level, 828 ONGs exhibiting PAVs in *O. sativa*, *O. glaberrima*, and *O. barthii*, and with available relative insertion positions were selected. A total of 74 ONGs that could not be assessed in the whole-genome resequencing (WGS) data were filtered out, leaving 754 ONGs for further analysis. Then, the analysis was conducted using sequencing data from 3458 accessions (Supplemental Table S4).

Multidimensional scaling (MDS) analysis of ONGs distinguishes well the Asian and African populations (Fig. 4A), *O. glaberrima* and *O. barthii* (Fig. 4B), four major groups of *O. sativa* (Fig. 4C), and three subpopulations of the *japonica* group (Fig. 4D) except for *indica* groups (Fig. 4E). PAV analysis of ONGs showed that there is considerable variation among and within the subpopulations of *O. sativa* (Fig. 4F; Supplemental Table S5). Among these ONGs, 422

and 432 ONGs showed different distributions among the four major *O. sativa* groups and the 15 subpopulations, respectively (Fisher's exact test P -value < 0.01). As shown in Figure 4F, 92 subpopulation-predominant, 183 major group-predominant, 29 *O. sativa*-predominant, 83 *O. glaberrima*-predominant, 42 *O. barthii*-predominant, and 142 *O. glaberrima/O. barthii*-predominant ONGs were observed. These results indicate that the distribution of ONGs is population and species specific. On this basis, we set out to find the representative ONGs in each major group and obtained three *Xian-indica* (XI) representative ONGs (ONG0900650, ONG0301080, and ONG1000640), one *Geng-japonica* (GJ) representative ONG (ONG0300140), one *circum-Aus* (cA) representative ONG (ONG0500890), and one *circum-Basmati* (cB) representative ONG (ONG0500930).

DNA-mediated organellar gene transfer

Experimental evidence has indicated that mtDNA escape is independent of RNA intermediates in yeast (Shafer et al. 1999), and plastid-to-nucleus transfer of DNA in plants can occur without the elimination of introns (Fuentes et al. 2012) or RNA editing sites

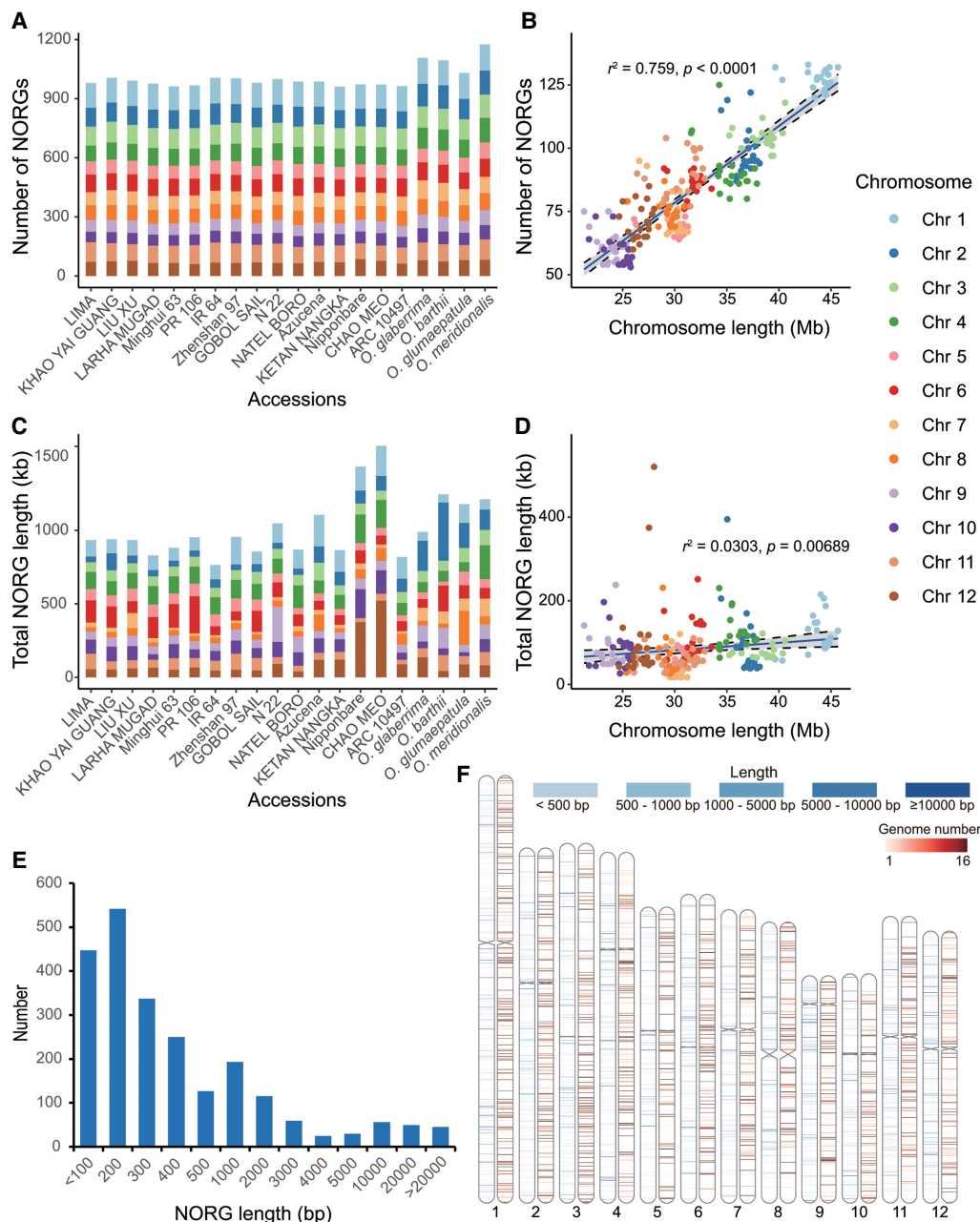


Figure 1. NORG length varies, but number and distribution are relatively constant across *Oryza* genomes. (A) NORG number per chromosome in each accession. (B) Correlation of NORG number with the chromosome size. (C) Total length of NORGs per chromosome in each accession. The total length of all NORGs from each chromosome was summed. (D) Sum of NORG lengths per chromosome in each accession correlated to chromosome size. (E) NORG length distribution. The x-axis indicates the length range of NORGs, with 200 indicating a length range of 100–200 bp, 300 indicating a length range of 200–300 bp, and so on. (F) Chromosomal distribution of NORGs in 16 *O. sativa* accessions. Chromosome heatmaps in the left and right represent the length and frequency of NORGs, respectively. A total of 1456 NORGs in 16 *O. sativa* assemblies were mapped to the *O. sativa* ssp. *japonica* cv. Nipponbare genome IRGSP-1.0.

(Sheppard et al. 2011). Further, an RNA-mediated model posits that mitochondrial genes could undergo retroprocessing in the mitogenome before being physically transferred as DNA (Wu et al. 2017). We examined the transfer of 15 organellar intron-containing genes and five *trans*-splicing genes and found no postsPLICING forms among NORGs. Only 11 NORGs contained one to four RNA-editing sites that had already changed to T bases, with no NORGs showing C-to-T change at all RNA editing sites

(Supplemental Table S6). These findings support that NORGs are transferred predominantly via DNA rather than via RNA or cDNA intermediates.

Signatures of junction sites of de novo NORG insertions

To better understand the mechanisms contributing to NORG formation, we examined the junction sequences of NORG insertions

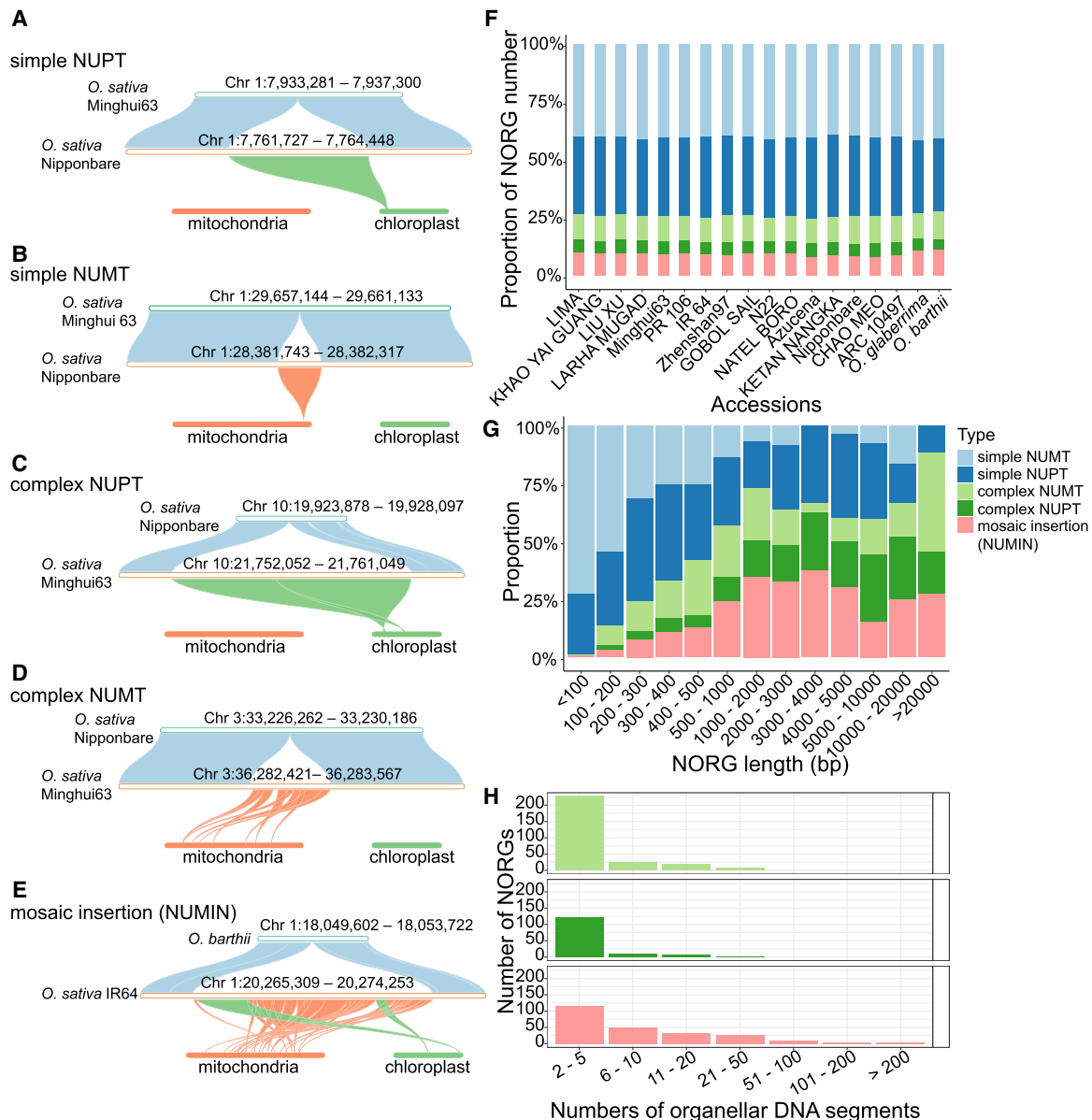


Figure 2. NORG types vary depending on length but are consistent across *Oryza* genomes. (A–E) Structure of five types of NORGs. (A) A simple NUPT consisting of one segment from plastid. Synteny between Minghui63 Chr 1: 7,933,281–7,937,300, ONG0100220 in Nipponbare, and the plastid of rice. (B) A simple NUMT consisting of one segment from mitochondria. Synteny between Minghui63 Chr 1: 29,657,144–29,661,133, ONG0100760 in Nipponbare, and the mitochondria of rice. (C) A complex NUPT consisting of two segments from the plastid. Synteny between Nipponbare Chr 10: 19,923,878–19,928,097, ONG1000620 in Minghui63, and the plastid of rice. (D) A complex NUMT consisting of eight segments from the rice mitochondria. Synteny between Nipponbare Chr 3: 33,226,262–33,230,186, ONG0301080 in Minghui63, and the mitochondria of rice. (E) A nuclear complex insertion (NUMIN) consisting of two segments from plastid and 26 segments from the mitochondria. Synteny comparison between *O. barthii* Chr 1: 18,049,602–18,053,722, ONG0100500 in *O. sativa* IR64, and the mitochondria and the plastid of rice. (F) Relative proportion of five types of NORGs in each accession. (G) Relative proportion of each type of NORGs with different lengths. The length of NORGs was calculated using the median size. (H) Number of organelle DNA segments in complex NORGs.

in 828 polymorphic ONGs from *O. sativa*, *O. glaberrima*, and *O. barthii*, for which we could obtain exact organelle DNA insertion sites. Microhomologies, blunt ends, and fillers were observed in the breakpoints (Fig. 5A; Supplemental Table S8), similar to the characteristics of T-DNA insertion (Kleinboelting et al. 2015). Microhomology (≥ 1 bp) was found in 48% of the NORG junctions, with the length of microhomology significantly exceeding the length expected by chance (chi-square test P -value

$< 2.2 \times 10^{-16}$) (Fig. 5A; Supplemental Fig. S7A). Moreover, target site duplication (TSD) and DNA loss were also found in NORG insertion sites (Fig. 5B). Among the 110 ONGs (13%) with TSD lengths > 5 bp, 36 had TSD > 20 bp, which is significantly longer than the length expected by chance (chi-square test P -value $< 2.2 \times 10^{-16}$) (Fig. 5A,D–E; Supplemental Table S8; Supplemental Fig. S7B). The majority of ONGs (73%, 607/828) appear to have been formed through DSB, with the primary mechanism being

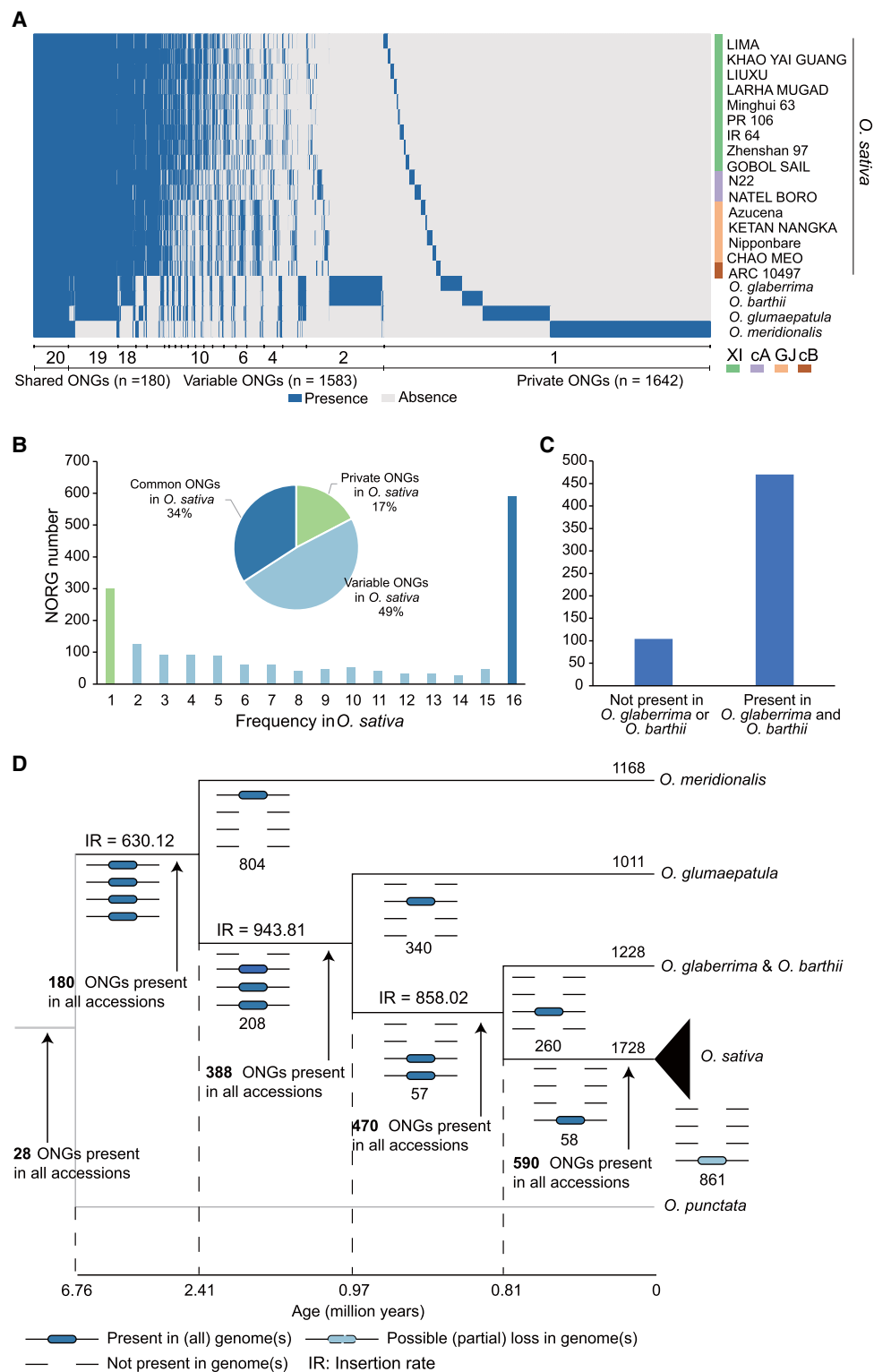


Figure 3. Abundant presence/absence variations of ONGs in *Oryza* genomes. (A) Presence and absence information of ONGs in five AA genome species. The presence of a NORG is indicated by blue, and the absence of a NORG is indicated by white. (B) Distribution of ONGs among 16 *O. sativa* accessions. The “common ONGs” are those present in all accessions; “variable ONGs” are present in more than one but fewer than 16 accessions; “private ONGs” are present in only one accession. (C) Distribution of *O. sativa* “common ONGs” in *O. glaberrima* and *O. barthii*. (D) Position of ONGs on the phylogenetic tree of AA genome species in *Oryza*. The total number of ONGs in genomes is shown above branches. The number of branch-specific ONGs is shown below branches. The four lines under each node from top to bottom represent *O. meridionalis*, *O. glumaepatula*, *O. glaberrima* and *O. barthii*, and *O. sativa*.

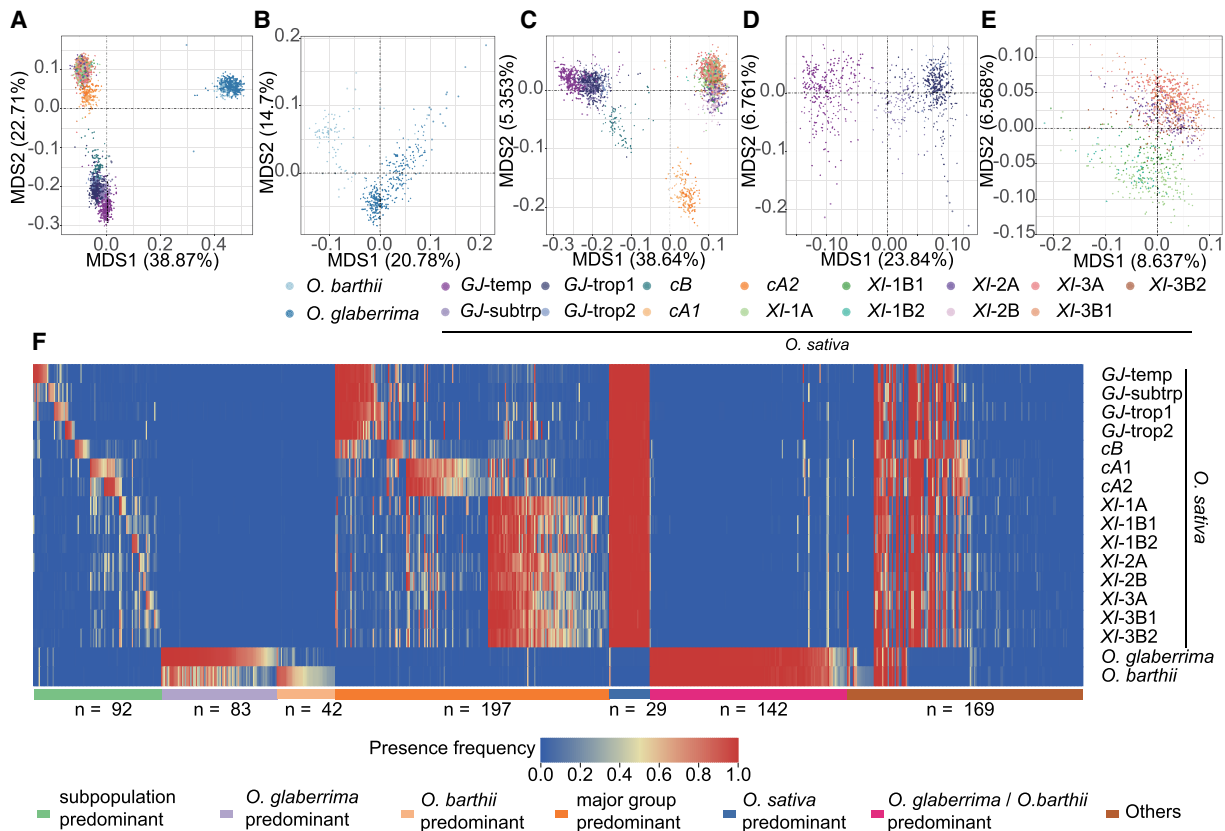


Figure 4. PAVs of NORGs reviews population structure and subpopulation differences. (cA) *circum-Aus*; (cB) *circum-Basmati*; (GJ) *Geng-japonica*, for which (trop) tropical, (subtrp) subtropical, and (temp) temperate; and (XI) *Xian-indica*. (A,E) Multidimensional scaling plots for all accessions of *O. sativa*, *O. glaberrima*, and *O. barthii* (A); *O. sativa*, *O. glaberrima* and *O. barthii* (C); GJ groups of *O. sativa* (D); XI groups of *O. sativa* (E). (F) Presence frequencies of NORGs (n = 754) plotted by heatmap.

classical non-homologous end joining (c-NHEJ; 59%, 975/1656) (Supplemental Table S8). Among these, 29% (179/607) of the NORGs exhibit distinct formation mechanisms at their two junction sites, with one being c-NHEJ and the other being alternative end joining (alt-EJ) or microhomology-mediated template switching (Fig. 5C; Supplemental Table S8). Fork stalling and template switching/microhomology-mediated break-induced replication (FoSTes/MMBIR) contributed to the formation of the remaining NORGs (27%, 221/828), resulting in both sequence loss and flanking sequence duplication that could not be explained by DSBR mechanisms (Supplemental Table S8). Collectively, these results indicate that both DSBR and replication-based mechanisms (RBMs) can generate NORGs.

Formation of complex NORGs

It remains to be determined whether complex NORGs evolved from simple NORG insertions or from more complex/aberrant recombination mechanisms. To address this, we propose three potential formation pathways: (1) single mutational events involving ligation of organelle DNA fragments before or during nuclear insertion (Supplemental Fig. S8A), (2) multiple insertion events at the same site (Supplemental Fig. S8B), and (3) postinsertion rearrangements after organelle DNA insertion (Supplemental Fig. S8C). First of all, methodologically, multiple insertions occurring in nearby regions were treated as independent NORGs, which ensures that complex NORGs formed by multiple insertion events

are not categorized as single NORG. In the 828 NORGs mentioned above, 33% (272/828) of the NORGs are complex NORGs (120 NUMTs, 51 NUPTs, and 101 NUMINs). Collinearity analysis of these NORGs and their flanking sequences indicated that these NORGs exhibit PAVs across different accessions, but there was no rearrangement after insertion. The insertion sites of simple and complex NORGs both suggested their formation through RBM and DSBR. Six conserved complex NORGs in *O. sativa* accessions were further examined. Collinearity analysis revealed these NORGs originated in and fixed in *O. sativa* without no sequential insertion or structural rearrangements postinsertion, strongly supporting their origin from a single event (Supplemental Fig. S9). In addition, 10 out of the 828 NORGs are found to be located within 5 kb regions around existing NORGs (Fig. 5F; Supplemental Fig. S10). Although rare cases of hypotheses 3 cannot be entirely excluded, systematic collinearity classification and structural analyses indicate that most complex NORGs arise via single mutational events.

Moreover, newly inserted NORGs can also be classified as complex NORGs (Supplemental Fig. S10A,C,G,H). For example, the simple NUPT ONG0200610 was formed in the common ancestor of *Oryza glumaepatula* and *O. sativa* (Fig. 5F). Subsequently, in the *O. sativa* clade, both new complex NORGs and simple NORGs inserted into its flanking region, with the most recently formed NORG (ONG020136) exhibiting the most complex form (Fig. 5F). These results suggest that both simple and complex NORGs can form regardless of the insertion time or rearrangement after insertion.

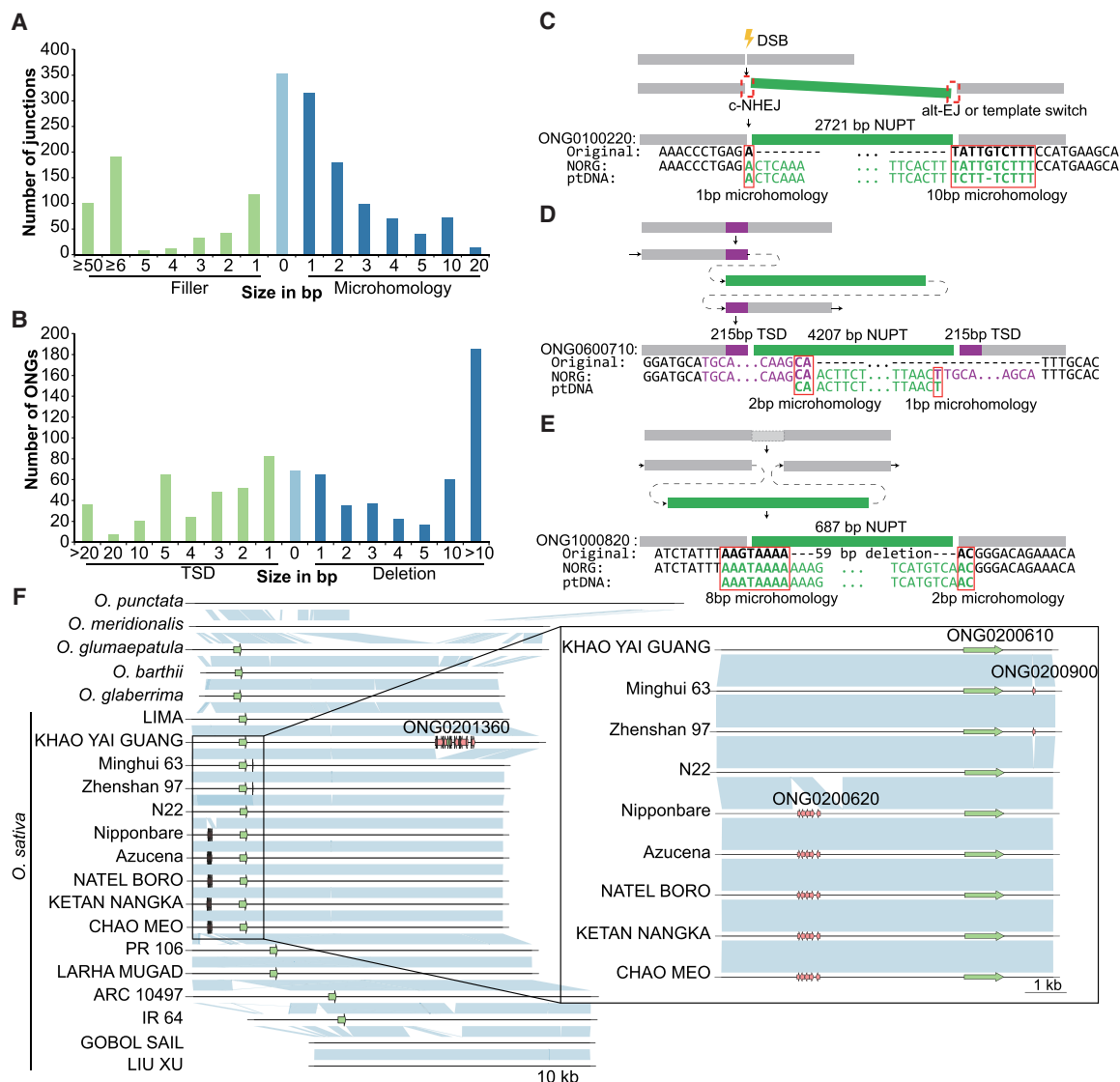


Figure 5. Characteristics of organelle DNA insertions reflecting its formation mechanisms. (A) Size distribution of microhomology and filler DNA at insertion sites of NORGs. (B) Size distribution of target-site duplications and deletions at insertion sites of NORGs. (C–E) Nucleotide-resolution breakpoint sequence of organelle DNA nuclear insertions. Red rectangles highlight sequence microhomology. (F) A region on Chr 2 has four independent NORGs among AA genome accessions.

In addition, among the 828 ORGs mentioned above, 11% (88/828) contained long DNA fillers >50 bp in the insertion sites, suggesting that NORG insertion can occur in conjunction with the insertion of nonorganelle DNA during single transfer events, leading to complex structural variations (Supplemental Table S8). In conclusion, complex NORGs could be formed by single mutational events, and the newly inserted NORGs near existing NORGs may also increase the complexity of NORGs.

NORG fragmentation triggered by transposable elements

Collinearity analysis revealed sequence insertions in 71 ORGs in one or more accessions, which resulted in the splitting of single NORGs into multiple fragments with very large distances between each other. The insertions were mainly composed of transposable elements (TEs), mostly the Ty3/Gypsy LTR retrotransposons

(Supplemental Fig. S11). Among the transposon insertion polymorphisms in rice, Ty3/Gypsy elements are most abundant (Huang et al. 2008). It has also been observed that the content of Ty3/Gypsy is particularly high in centromeric regions compared with in other regions (Song et al. 2021). Once a NORG is integrated into nuclear genome, it becomes part of the nuclear DNA and, like other sequences, is subjected to the influence of TEs. Thus, the presence of Ty3/Gypsy insertions within NORGs may reflect their general abundance and activity within the genome, rather than a NORG-specific association.

Based on their original and derived forms, we further interpreted the dynamic change process of NORGs. For example, ONG0400870 appears to have been a 66 kb insertion of mtDNA along with >10 kb sequences from chromosomal DNA in the ancestor of the AA genome, which underwent independent changes (i.e., sequence loss and TE insertion) in different clades (Fig. 6A,B).

After speciation, seven insertions occurred in *Oryza meridionalis*, whereas an independent insertion and large segment loss occurred in *O. glumaepatula*. In the common ancestor of *O. sativa*, *O. glaberrima*, and *O. barthii*, a segment loss occurred in the middle of the NUMT. In *O. sativa*, several independent insertions occurred within NORGs, resulting in their dispersion and fragmentation (Fig. 6A,B). Similarly, ONG0400650 formed in the common ancestor of *O. sativa*, which contains two separate NUPTs. The second NUPT then underwent several independent TE insertions in different accessions (Fig. 6C,D). Multiple changes from transposons were likewise found in ONG0100920, ONG0900660, and others (Supplemental Fig. S12).

Discussion

In this study, a high-throughput identification pipeline for NORGs was established and used to reveal the landscape of NORGs in *Oryza* using 16 *O. sativa* genomes representative of all 15 subpopulations and six other wild relatives of *Oryza*. Our methods based on collinearity between different accessions were used to identify and analyze ONGs, thereby providing a more accurate detection process of organelle DNA insertion and change, which is similar to the approach and advantages described by Uvizl et al. (2024). NORGs can be fragmented by insertions into multiple segments separated at distances of several to tens of kilobases (Fig. 6). At the same time, there are also cases in which new NORGs are inserted within 5 kb of pre-existing NORGs (Fig. 5F). Such changes cannot be identified by determining whether the adjacent sequences from organelle DNA are the same NORGs according to a threshold (such as 5 kb intervals).

Importantly, we found that the predominant forms of NORGs in *Oryza* were mostly <1000 bp in length (Fig. 1E), and longer NORGs were mainly complex NORGs (Fig. 1G). Few intact and contiguous long organelle DNA fragments were observed. Similarly, the majority of NUMTs in humans (Wei et al. 2022), as well as the NUMTs and NUPTs in *Triticum/Aegilops* complex species (Zhang et al. 2024b) and *Toxoplasma* (Namasivayam et al. 2023), are short insertions, indicating that the short-fragmented nature of NORGs is widespread in eukaryotes. Small fragments of NUMT can be inserted into the site of DSB in tobacco, which may result from extensive fragmentation of mtDNA (Wang et al. 2012). Previous studies have found that microinjected plasmid DNA in HeLa and COS cells disappeared within 50–90 min owing to the rapid plasmid degradation caused by cytosolic nuclease(s) (Lechardeur et al. 1999). In addition, the rate of DNA diffusion through the cytoplasm is heavily dependent on the DNA size (Lukacs et al. 2000), and therefore, small fragments of free organelle DNA are more likely to enter the nucleus. In summary, we suspect that damage or fragmentation of free organelle DNA in the nucleus may be the primary reason for the presence of more short-fragment NORGs than long-fragment NORGs.

The escape of DNA from chloroplast to nucleus is estimated to occur once per 16,000 pollen grains (Huang et al. 2003) and once per 5 million somatic cells (Stegemann et al. 2003). Such a high frequency has also been observed in mitochondria, with a rate of approximately 2×10^{-5} per cell per generation in yeast (Thorsness and Fox 1990) and once in every 4000 births in humans (Wei et al. 2022). This work estimates the rate of organelle-to-nucleus DNA transfer (and retention) at an evolutionary scale in plants. In this study, it was estimated that the insertion rate of ONGs in AA genome was 631.12–943.81 insertions per million years (Fig. 3D), suggesting that there has been a significant amount of de

novo insertions of organelle DNA in the nucleus genome during the evolution of *Oryza* and that this process is ongoing and frequent.

In addition, the evolutionary branches with greater distances tended to have fewer conserved ONGs, indicating that some NORGs may be gradually lost owing to nuclear genomic variations during evolution (Fig. 3D). Once formed, NORGs will be subjected to persistent mutations in the nuclear genome, such as base mutations and insertions and deletions (indels) (Huang et al. 2005), and may be lost in progeny owing to their instability (Sheppard and Timmis 2009). Here, we found that NORGs, especially nonconserved ones, are primarily located in intergenic regions (Supplemental Fig. S6; Supplemental Table S3), suggesting selection against insertions that may disrupt genes. In *Triticum/Aegilops* complex species, NUPTs and NUMTs also tend to be located in intergenic regions, as accompanied by high levels of DNA methylation and low levels of active histone modification (Zhang et al. 2024b). These findings collectively suggest that NORGs primarily constitute nonfunctional sequences, which may be the reason for the low conservation of NORGs observed in this study. Moreover, we suspect that TE activity is an important cause of NORG fragmentation and diverse variations in different accessions (Fig. 6; Supplemental Fig. S12). Under the action of mutations, sequence loss, and TE activity, NORGs may be lost or become no longer detectable, which may also lead to the nonconservation nature of NORGs. These factors together contribute to the rich polymorphisms of NORGs together with continuous and frequent NORG insertions, which have continuously reshaped the *Oryza* nuclear genome.

Unlike conservation of mitochondrial gene content in metazoa (Castellana et al. 2011; Shtolz and Mishmar 2023), the loss of mitochondrial and plastid genes in angiosperms is ongoing (Mohanta et al. 2020; Mower 2020). Several organellar genes, such as the mitochondrial ribosomal protein genes (Adams et al. 2000, 2002) and the plastid acetyl-CoA carboxylase subunit D (*accD*) gene (Harris et al. 2013; Rousseau-Gueutin et al. 2013), have been lost and transferred to the nucleus multiple times during angiosperm evolution. There are abundant copies of organellar genes in the nuclear genome of each accession (Supplemental Tables S6, S7). Among the 180 shared ONGs in AA genomes, 14 contain NIOGs. Only two of these ONGs (ONG0600060 and ONG0700470), which include *ORF241* (OrsajM_p39) and OrsajCt151, were also found in *O. punctata* (BB genome) (Supplemental Table S6). These observations suggest that NIOGs are continuously formed as organellar DNA is inserted into the nuclear genome, and they lack conservation during evolution, which is similar to the findings of Chen et al. (2023). The expression patterns of plant organelle genes are different from those of nuclear genes (Liere et al. 2011; Zhang et al. 2023). To become functional, NIOGs typically require acquisition of a nuclear promoter, a polyadenylation signal for expression, and, if necessary, signal peptides or alternative mechanisms to target the protein back to the mitochondria or plastid (Stegemann and Bock 2006; Liu et al. 2009). Unlike parasitic genetic elements, DNA from organelles might insert into the regulation regions, such as promoters, untranslated regions, and introns, or land in the nuclear genome with intact reading frames or protein domains, which help contribute the raw genetic material needed for gene expression regulation and de novo gene birth.

Our data demonstrated that inter- and intra-population polymorphism of ONGs and PAVs of ONGs can be utilized for reliable clustering of rice populations (Fig. 4A–E). Most of these ONGs are

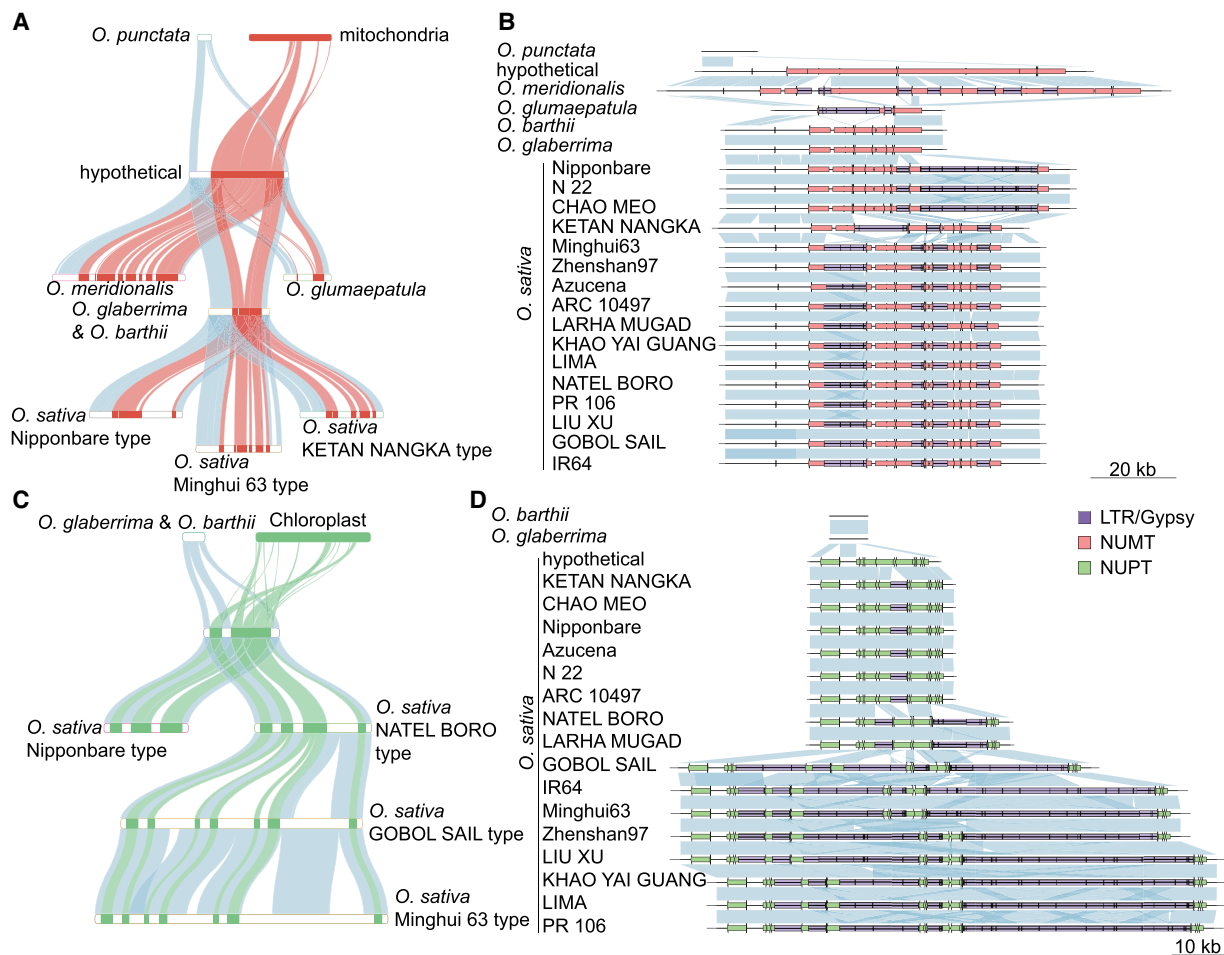


Figure 6. Transposons result in the fragmentation of NORGs. (A) Dynamic process of ONG0400870. The “hypothetical” indicates the insertion form of ONG0400870 in the presumed ancestor of the AA genome, which is inferred from the form of ONG0400870 in the AA genome. The orange color indicates the mtDNA sequence. (B) Synteny of the ONG0400870 region in the *O. punctata* (BB genome) and AA genomes. The NORG is not present in *O. punctata*. Orange indicates the NUMT. (C) Dynamics of ONG0400650. The “hypothetical” indicates the insertion form of ONG0400650 in the presumed ancestor of the *O. sativa* genomes, which is inferred from the form of ONG0400650 in *O. sativa*. Green color indicates ptDNA sequence. (D) Synteny of the ONG0400650 region in *O. barthii*, *O. glaberrima*, and *O. sativa*. The NORG is not present in *O. barthii* and *O. glaberrima*. Green color indicates NUPTs.

shared by at least two subpopulations ($K = 15$) in *O. sativa* (Fig. 4F), indicating some admixture among subpopulations during the domestication and breeding of *O. sativa*. As NORGs are homoplasmy-free and easily identified (Hazkani-Covo 2009; Liang et al. 2018), they will be valuable tools for phylogenetic studies.

Direct transfer of organellar DNA into the nuclear genome and transfer via mRNA/cDNA intermediates have been considered as the two main mechanisms for endosymbiotic gene transfer. We did not observe gene transfer that involves the removal of introns in NORGs (Supplemental Table S6). Because of the presence of mutations and U encoding at the DNA level, the few RNA-edited sites in NIOGs (Supplemental Table S6) may not result from RNA- or cDNA-mediated gene transfer. These findings collectively suggest that the organellar gene transfer to the nuclear genome predominantly occurs via DNA rather than via RNA or cDNA intermediates, although the possibility of RNA-mediated transfer cannot be entirely excluded.

Previous studies have revealed that organelle DNA can be captured during DSBR (Ricchetti et al. 1999; Yu and Gabriel 1999; Decottignies 2005; Wang et al. 2012, 2018a), but the molecular mechanism of organelle DNA insertion into the nucleus has not

been fully elucidated. DNA double-strand break (DSB) can be repaired by c-NHEJ (involving blunt end and short microhomologies from 1 to 4 bp), alt-EJ (involving microhomologies <20 bp), homologous recombination (HR; typically requires >50 bp homology) and single-strand annealing (SSA; typically requires >20 bp homology) (Chang et al. 2017). In *Toxoplasma gondii*, the capture of ptDNA (from apicoplast, a type of plastid) and mtDNA at DSB sites is dependent on Ku80 protein, which plays a crucial role in the c-NHEJ pathway of DNA repair, suggesting that DSBR via the c-NHEJ pathway is the key mechanism for the formation of NORGs in *T. gondii* (Namasivayam et al. 2023). In this study, only microhomologies <20 bp were found (Fig. 5A), indicating that the insertion of NORGs does not depend on HR or SSA throughout *Oryza* evolution. In addition, we found that alt-EJ or microhomology-mediated template switching could also contribute to integration of organelle DNA. Consistent with this, a previous study revealed that DNA polymerase theta-mediated DNA end joining (TMEJ; the predominant mediator of alt-EJ in most eukaryotes) promotes random integration of exogenous DNA in mouse embryonic stem cells (Zelensky et al. 2017). Moreover, long TSDs (>20 bp) and deletions (>10 bp) were observed in the insertion

site of some NORGs (Fig. 5D,E; Supplemental Table S8), which are difficult to explain by c-NHEJ- or alt-EJ-mediated mechanisms. In addition to DSBR pathways, RBMs such as FoSTeS/MMBIR could explain nonrecurrent structural variants such as complex genomic rearrangements containing multiple breakpoints, insertions of DNA segments, and microhomology at the breakpoint junctions (Lee et al. 2007; Bursted et al. 2022). These mechanisms can result in complex rearrangements in genomic disorders (Lee et al. 2007; Zhang et al. 2009), template insertions in cancer cells (Osia et al. 2021), and structural variants in rice (Qin et al. 2021). In human cancer cells, somatic mtDNA integration events frequently occur together with other rearrangements of the nuclear genome (Ju et al. 2015), suggesting that mtDNA may be used as a “filler material” for DNA replication through MMBIR (Ju et al. 2015; Ju 2016). Similarly, ptDNA and mtDNA may also act as DNA fillers and be captured during DNA repair or DNA replication. Therefore, we hypothesize that NORGs result from DSBR and replication errors without requiring long homologies.

Based on this study, we propose that the formation of complex NORGs occurs before or during the process of insertion via ligation of ptDNA and/or mtDNA rather than by subsequent insertion or rearrangement after insertion. In addition, insertion of NORGs can be accompanied by filling with nonorganelle DNA, resembling the insertion pattern of mtDNA in human somatic cells (Ju et al. 2015). Similarly, particle bombardment in transgenic papaya SunUp resulted in insertion of a 1.6 Mb foreign DNA sequence containing several newly formed NUPTs, NUMTs, and sequences from the nuclear genome (Yue et al. 2022). DSBR could also result in complex insertions originating from various regions on different chromosomes via multiple sequential invasion and annealing events (Min et al. 2023). Because of possible damage or fragmentation of organelle DNA entering the nucleus, template switching may be further triggered during integration of organelle DNA, forming complex NORGs containing several to hundreds of fragments through mechanisms such as FoSTeS/MMBIR. Additionally, new organelle DNA insertions can accidentally occur near the existing NORGs owing to continuous formation of NORGs, which further increases the complexity of NORG structure.

Our study relies on a single mitochondrial reference genome (*O. sativa*) for NUMT identification, necessitated by the lack of published mtDNA assemblies for other *Oryza* species. Although mitochondrial sequence composition is highly conserved across the genus, structural variations such as recombination or lineage-specific rearrangements may result in undetected NUMTs. Alignment-based methods like BLAST are particularly susceptible to missing NUMTs derived from divergent mtDNA regions, as they prioritize sequence continuity over fragmented homology. Recent advances in alignment-free approaches (Li et al. 2019; for review, see Xue et al. 2023) offer promising solutions, which can detect NUMTs even when structural divergence obscures sequence collinearity. Future applications of such *k*-mer-based analyses may further resolve species-specific NUMTs and provide a more comprehensive understanding of NORG diversity in *Oryza*.

In summary, we classified NORGs into ONGs and analyzed their characteristics and polymorphisms in *Oryza* at the pangenome level. Our findings reveal continuous insertion and nonconservation of NORGs during evolution, resulting in intra- and interspecies variations. Our results indicate that organelle DNA might be integrated into the nucleus through DSBR and replication-based template switch, and complex NORGs may arise from single insertion events. The ongoing insertions and losses of NORGs,

coupled with TE-induced fragmentation, collectively contribute to the polymorphism of NORGs throughout *Oryza* evolution. All these findings provide new insights into the ongoing and dynamic evolutionary processes of intracellular DNA promiscuity in plants.

Methods

Genome resources

A total of 22 genomes covering AA, BB, and FF genome species in the *Oryza* genus were collected, including 16 high-quality genome assemblies covering 15 subpopulations of Asian cultivated rice *O. sativa* (Kawahara et al. 2013; Zhou et al. 2020; Song et al. 2021) and six genome assemblies of other species (*O. glaberrima*, *O. barthii*, *O. glumaepatula*, *O. meridionalis*, *O. punctata*, and *Oryza brachyantha*) (Stein et al. 2018; Ntakirutimana et al. 2023; Thathapalli Prakash et al. 2023). The genome sequences in FASTA format were obtained for each assembly. The genome sequences and gene annotations of *O. sativa* Nipponbare were downloaded from Ensembl Plants (<http://plants.ensembl.org/>). The genome sequences and gene annotations of other *O. sativa* accessions were downloaded from RGI (Yu et al. 2023; <https://riceome.hzau.edu.cn>), and the assembly data of CHAO MEO::IRGC 802731 (GCA_009831315.1), Azucena (GCA_009830595.1), KETAN NAN GKA::IRGC 19961-2 (GCA_009831275.1), ARC 10497::IRGC 12485-1 (GCA_009831255.1), PR 106::IRGC 53418-1 (GCA_009831045.1), Minghui 63 (GCA_001623365.2), IR 64 (GCA_0098314875.1), Zhenshan 97 (GCA_001623345.2), LIMA::IRGC 81487-1 (GCA_009829395.1), KHAO YAI GUANG::IRGC 65972-1 (GCA_009831295.1), GOBOL SAIL (BALAM)::IRGC 26624-2 (GCA_009831025.1), LIU XU::IRGC 109232-1 (GCA_009829375.1), LARHA MUGAD::IRGC 52339-1 (GCA_009831355.1), N22 (N 22::IRGC 19379-1; GCA_001952365.2), and NATEL BORO::IRGC 34749-1 (GCA_009831335.1), plus the remaining assemblies of *O. glaberrima*::IRGC 96717 (GCA_000147395.3), *O. barthii*::IRGC 105608 (GCA_000182155.4), *O. glumaepatula* (GCA_000576495.2), *O. meridionalis*::OR44 (W2112; GCA_000338895.3), *O. punctata*::IRGC 105690 (GCA_000573905.2), and *O. brachyantha*::IRGC 101232 (GCA_000231095.3) were downloaded from NCBI's Genome database (<https://www.ncbi.nlm.nih.gov/datasets/genome/>).

Identification of ONGs

We used the *O. sativa* chloroplast genome to identify NUPTs owing to the chloroplast genome is highly conserved among the species used in this study. The mitochondrial genome of *O. sativa* (NCBI GenBank [<https://www.ncbi.nlm.nih.gov/genbank/>] NC_011033) was used as the reference for NUMT identification. This choice was necessitated by the lack of published mitochondrial assemblies for other *Oryza* species included in our analysis. Despite the structural diversity of plant mitochondrial genomes, their sequence composition is highly conserved across species. To validate the suitability of the *O. sativa* mitochondrial genome as a universal reference, we compared it to the mitochondrial assembly of *Oryza minuta* (BBCC genome, NCBI GenBank: NC_029816.1), a closely related species, using whole-genome alignment (NUCmer). The comparison revealed 97% sequence identity and 83.1% coverage, supporting the use of *O. sativa* mtDNA for NUMT detection in closely related *Oryza* species.

We used NUCmer (Marçais et al. 2018) for sequence alignment (version 4.0.0beta2, parameters: -c 50 --maxmatch). NC_001320 (NCBI GenBank) was used for plastid genome, and NC_011033 (NCBI GenBank) was used for mitochondrial genome as the query sequence, respectively. All anchor matches, regardless

of their uniqueness, were used (parameter: --maxmatch), and the minimum length of a cluster of matches was set to 50 (parameter: -c 50). These parameters allow the inclusion of shorter matches and retain more matches than the default settings. The overlapping hits were merged by the BEDTools (Quinlan and Hall 2010) merge command, and the results were utilized for calculating the total length of the NORGs. Then, the alignments separated by <5 kb of DNA of nonorganelle origin were merged by BEDTools merge command, and the results were temporarily used as NORG clusters.

Each NORG cluster including its flanking sequences of 300 bp was aligned to all genomes using NUCmer with the default parameters, followed by show-coords (parameters: -rclT -O). First, the locations of the NORG clusters in other genomes were determined by the .coords file generated by show-coords. If the genome contained the same NORG cluster without structure variations, the alignment was marked as “CONTAINS” in the .coords file. For the remaining NORG clusters, the best location of each NORG cluster on each genome was determined by show-tiling (parameters: -g 20000). The NORG clusters in the same locations were recognized as the same NORGs by BEDTools intersect.

Because some NORGs were separated by sequence insertions >5 kb, multiple NORG clusters at the same location were merged. In the first step, the alignment of NORGs with flanking sequence of 300 bp from the previous step was used directly to merge partially dispersed NORGs. The dot plots of each NORG cluster produced by mummerplot were used as an aid to manual calibration. In the second step, the flanking sequences of 2000 bp from each NORG cluster in step one were aligned with the genomes without the identification of that NORG. The .delta file from the alignment was then treated with show-tiling (parameters: -g 20000 -v 80) to determine the position of the flanking sequences in each genome. In the third step, several NORGs with multiple large insertions were corrected manually using BLASTN (Camacho et al. 2009) and BLAT (Kent 2002). To assess the accuracy of this workflow, synteny plots of each ONG were generated by genoPlotR (Guy et al. 2010) to visually validate the NORGs.

Identification of NORG insertion position

For each NORG, the flanking 2 kb sequences were merged and aligned with the genomes in which the NORG was not detected using NUCmer, after which the positions of the flanking 2 kb merged sequences in other genomes were obtained using show-tiling (parameters: -g 20000 -v 80), and the NORGs whose flanking sequences could not be obtained in a relative position by show-tiling were ignored. Afterward, the R package genoPlotR was used to plot the collinearity of all NORGs and their flanking sequences in all genomes in which NORGs and insertion positions could be detected, which was used to assist manual correction.

Statistical analysis of NORG overlap with gene and exon regions

The gene annotations of *O. sativa* Nipponbare were downloaded from Ensembl Plants (<http://plants.ensembl.org/>). The gene annotations of *O. sativa* Minghui 63 and Zhenshan 97 were downloaded from Rice Information GateWay (RIGW) (Song et al. 2018; http://rice.hzau.edu.cn/rice_rs3/). The gene annotations of other *O. sativa* accessions were downloaded from RGI (Yu et al. 2023; <https://riceome.hzau.edu.cn>). To assess whether the observed overlap between NORGs and gene or exon regions was statistically significant, random permutation was performed using BEDTools. First, the “bedtools shuffle” command was employed to randomly permute the genomic locations of the identified NORGs while maintaining the original number and length of the NORGs. This

generated a set of simulated NORGs with randomized positions across the genome. This process was repeated 1000 times to generate a distribution of expected overlaps. Next, the “bedtools intersect” command was used to count the number of simulated NORGs that overlapped with gene regions and exon regions in each permutation, which could provide the expected values for the overlap. Similarly, “bedtools intersect” was used to determine the actual number of NORGs overlapping with gene regions and exon regions.

For each permutation, a chi-square test was performed to compare the observed overlaps with the expected overlaps using the “CHISQ.TEST” function in Excel. The *P*-value was calculated for each permutation, and then, the average *P*-value across all 1000 permutations was taken to assess the statistical significance of the observed overlaps.

NORG insertion rate estimations

Similar to the method of Zhou et al. (2023), to estimate the NORG insertion rate (NIR), we divided the total number of ONGs in pairs of genomes by twice the time to the most recent common ancestor (TMRCA). The estimate is calculated using the following equation:

$$\text{NIR} = \frac{\text{Number of ONGs}}{2 \times \text{TMRCA}} \quad (1)$$

In detail, the number of ONGs present only in *O. sativa* was 919. The number of ONGs present only in *O. glaberrima* and/or *O. barthii* was 471. The diverged time was 0.81 million years ago (MYA). The estimate was calculated using the following equation:

$$\text{NIR} = \frac{919 + 471}{2 \times 0.81} = 858.02.$$

The number of ONGs present in *O. sativa*, *O. glaberrima*, and/or *O. barthii* but not present in other genomes was 1491. The number of ONGs present only in *O. glumaepatula* was 340. The diverged time was 0.97 MYA. The estimate was calculated using the following equation:

$$\text{NIR} = \frac{1491 + 340}{2 \times 0.97} = 943.81.$$

The number of ONGs present in genomes other than *O. meridionalis* is 2238. The number of ONGs present only in *O. meridionalis* was 804. The diverged time was 2.41 MYA. The estimate is calculated using the following equation:

$$\text{NIR} = \frac{2238 + 804}{2 \times 2.41} = 630.12.$$

Characterization of NORG junction sites

To avoid errors caused by mutations and for more accurate characterization of the NORG junctions, we selected 18 genomes of *O. sativa*, *O. glaberrima*, and *O. barthii* for analysis. Through the insertion position of NORGs obtained previously, two genotype sequences of the presence and absence of NORGs were obtained. The sequences of the presence genotype of NORGs were the sequences of the NORG with its flanking 2 kb sequences, and the sequences of the absence genotype of NORGs were obtained by the position of the flanking 2 kb sequences of the NORGs in the genome that does not contain the NORGs. For each NORG, the sequences of both genotypes were aligned by BLASTN, and the HSPs flanking the NORG closest in distance to the NORG were retained, respectively. The number of overlapping base pairs of the closest HSP to the NORG was defined as microhomology (≤ 20 bp) or homology (>20 bp). The number of base pairs of the closest

HSP to the NORG that were >0 bp was defined as a sequence insertion. The closest HSP with exactly 0 bp distance from NORG was defined as a blunt end. The formation mechanisms were characterized as follows: classical NHEJ, >5 bp filler or <5 bp microhomology in the junction sites; alt-EJ, 5–20 bp microhomology in the junction sites; SSA, 21–50 bp homology in the junction sites; HR, >50 bp homology in the junction sites; and FoSTeS/MMBIR, >20 bp TSD or >10 bp deletion in the junction sites.

Determination of ONG presence or absence based on WGS data sets

The WGS data set was collected from 3K-RG (Wang et al. 2018b) and the studies of Cubry et al. (2018), Meyer et al. (2016), and Choi et al. (2019) from the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject/>) under accession numbers PRJEB6180, PRJEB21312, PRJNA315063, and PRJNA453903, respectively.

To identify the PAVs of ONGs in the rice WGS data set, we developed a pipeline based on presence/absence genotype across assemblies through four major steps: (1) sequences of presence/absence genotype of ONGs were prepared using the identification results described above; (2) to reduce the mapping errors caused by repetitive sequences and organelle genome sequences, the rice TE library “rice6.9.5.liban” from EDTA (Ou et al. 2019) and rice organellar genome NC_001320 (NCBI GenBank) and NC_011033 (NCBI GenBank) were merged to the sequences of the previous step; (3) paired-end reads were mapped to the merged sequences by BWA-MEM2 (Vasimuddin et al. 2019); and (4) the presence/absence genotype was verified by the mapping results.

To verify that this pipeline could accurately identify ONG presence/absence, we chose five WGS accessions (CX140 | IRGSP1.0, CX145 | MH63RS3, CX133 | ZS97RS3, CX145 | AzucenaRS1, IRIS_313-8813 | Os117425RS1) from 3K-RG corresponding to the genome assemblies used in this study. After filtering out the misidentified ONGs, 754 polymorphic NORGs were retained.

The MDS plots were constructed using the presence/absence information of 754 ONGs for each accession. The distance matrices were calculated using the “bray” method in the “vegdist” function from the R package “vegan” and subsequently computed the MDS using the “cmdscale” function in R (version 4.2.3) (R Core Team 2023).

The presence frequency of ONGs was calculated by using R software (version 4.2.3). The heatmap was plotted by TBtools software (Chen et al. 2020).

Identification of population specific ONGs

We defined the ONGs with presence frequency $\geq 90\%$ in one group and $\leq 10\%$ in other groups as representative ONGs among the four major groups. Considering that the same criteria could not be applied to the subpopulations, we lowered the subpopulation presence frequency threshold to 80%.

TE annotation

The TEs were identified by RepeatMasker v4.1.0 with parameter “-nolow -gff -xsmall -pa 4 -species *Oryza*” using the RepBase-20181026 and Dfam_3.1 library.

Visualization of ONGs

The schematics for five types of NORGs and the NORG change trajectories were plotted using NGenomeSyn (He et al. 2023). The synteny plots of NORGs were generated by genoPlotR (Guy et al.

2010). The R package RIdeogram (Hao et al. 2020) was used to visualize the location of ONGs in the chromosomes. The insertion frequency (depth) plots were generated by ggcoverage (Song and Wang 2023), and the genome maps of mtDNA and ptDNA were generated by OGDRAW (Greiner et al. 2019) and PMGmap (Zhang et al. 2024a), respectively.

Software availability

The source codes used in this study are available as Supplemental Code.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

This research was supported by the National Key R&D Program of China (2023ZD04073), the Foundation of Hubei Hongshan Laboratory (2022HSZD014), Major Project of Hubei Hongshan Laboratory (2022HSZD031), National Natural Science Foundation of China (31771752), the Fundamental Research Program of Hubei Province (2024AFE001), the Start-up Fund of the Huazhong Agricultural University (HZAU) to J.Z., and the Earmarked Fund for China Agriculture Research System (CARS-01-01). We thank the National Key Laboratory of Crop Genetic Improvement, HZAU for providing the bioinformatics computing platform in this study. We thank Dr. Laura Bartley for constructive criticism of the manuscript.

Author contributions: F.Z., J.Z., and Y.L. designed and conceived the research. C.G., Y.H., M.L., Y.Z., Y.X., N.M., X.Q., and A.Z. performed the research. R.A.W., J.Z., and Y.Z. contributed the materials. Y.H., M.L., Y.X., and X.Q. analyzed data. C.G., W.X., R.A.W., F.Z., J.Z., and Y.L. wrote and edited the paper. All authors read and approved the final manuscript.

References

- Adams KL, Daley DO, Qiu Y-L, Whelan J, Palmer JD. 2000. Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature* **408**: 354–357. doi:10.1038/35042567
- Adams KL, Qiu Y-L, Stoutemyer M, Palmer JD. 2002. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proc Natl Acad Sci* **99**: 9905–9912. doi:10.1073/pnas.042694899
- Bursedd B, Zamariolli M, Bellucco FT, Melaragno MI. 2022. Mechanisms of structural chromosomal rearrangement formation. *Mol Cytogenet* **15**: 23. doi:10.1186/s13039-022-00600-6
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421. doi:10.1186/1471-2105-10-421
- Castellana S, Vicario S, Saccone C. 2011. Evolutionary patterns of the mitochondrial genome in Metazoa: exploring the role of mutation and selection in mitochondrial protein-coding genes. *Genome Biol Evol* **3**: 1067–1079. doi:10.1093/gbe/evr040
- Chang HHY, Pannunzio NR, Adachi N, Lieber MR. 2017. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol* **18**: 495–506. doi:10.1038/nrm.2017.48
- Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant* **13**: 1194–1202. doi:10.1016/j.molp.2020.06.009
- Chen Y, Guo Y, Xie X, Wang Z, Miao L, Yang Z, Jiao Y, Xie C, Liu J, Hu Z, et al. 2023. Pangenome-based trajectories of intracellular gene transfers in Poaceae unveil high cumulation in Triticeae. *Plant Physiol* **193**: 578–594. doi:10.1093/plphys/kiad319
- Choi JY, Zaidem M, Gutaker R, Dorph K, Singh RK, Purugganan MD. 2019. The complex geography of domestication of the African rice *Oryza glaberrima*. *PLoS Genet* **15**: e1007414. doi:10.1371/journal.pgen.1007414

- Cubry P, Tranchant-Dubreuil C, Thuillet A-C, Monat C, Ndjiondjop M-N, Labadie K, Cruaud C, Engelen S, Scarcelli N, Rhoné B, et al. 2018. The rise and fall of African rice cultivation revealed by analysis of 246 new genomes. *Curr Biol* **28**: 2274–2282.e6. doi:10.1016/j.cub.2018.05.066
- Dayama G, Zhou W, Prado-Martinez J, Marques-Bonet T, Mills RE. 2020. Characterization of nuclear mitochondrial insertions in the whole genomes of primates. *NAR Genomics Bioinforma* **2**: lqaa089. doi:10.1093/nargab/lqaa089
- Decottignies A. 2005. Capture of extranuclear DNA at fission yeast double-strand breaks. *Genetics* **171**: 1535–1548. doi:10.1534/genetics.105.046144
- Ellis J. 1982. Promiscuous DNA—chloroplast genes inside plant mitochondria. *Nature* **299**: 678–679. doi:10.1038/299678a0
- Fuentes I, Karcher D, Bock R. 2012. Experimental reconstruction of the functional transfer of intron-containing plastid genes to the nucleus. *Curr Biol* **22**: 763–771. doi:10.1016/j.cub.2012.03.005
- Greiner S, Lehwerk P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res* **47**: W59–W64. doi:10.1093/nar/gkz238
- Gualberto JM, Newton KJ. 2017. Plant mitochondrial genomes: dynamics and mechanisms of mutation. *Annu Rev Plant Biol* **68**: 225–252. doi:10.1146/annurev-arplant-043015-112232
- Guo X, Ruan S, Hu W, Cai D, Fan L. 2008. Chloroplast DNA insertions into the nuclear genome of rice: the genes, sites and ages of insertion involved. *Funct Integr Genomics* **8**: 101–108. doi:10.1007/s10142-007-0067-2
- Guy L, Roat Kultima J, Andersson SGE. 2010. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* **26**: 2334–2335. doi:10.1093/bioinformatics/btq413
- Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G, Chen J. 2020. *Rldeogram*: drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput Sci* **6**: e251. doi:10.7717/peerj-cs.251
- Harris ME, Meyer G, Vandergon T, Vandergon VO. 2013. Loss of the acetyl-CoA carboxylase (accD) gene in Poales. *Plant Mol Biol Report* **31**: 21–31. doi:10.1007/s11105-012-0461-3
- Harutyunyan T. 2024. The known unknowns of mitochondrial carcinogenesis: *de novo* NUMTs and intercellular mitochondrial transfer. *Mutagenesis* **39**: 1–12. doi:10.1093/mutage/gead031
- Hazkani-Covo E. 2009. Mitochondrial insertions into primate nuclear genomes suggest the use of numts as a tool for phylogeny. *Mol Biol Evol* **26**: 2175–2179. doi:10.1093/molbev/msp131
- Hazkani-Covo E, Covo S. 2008. *Numt*-mediated double-strand break repair mitigates deletions during primate genome evolution. *PLoS Genet* **4**: e1000237. doi:10.1371/journal.pgen.1000237
- Hazkani-Covo E, Zeller RM, Martin W. 2010. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genet* **6**: e1000834. doi:10.1371/journal.pgen.1000834
- He W, Yang J, Jing Y, Xu L, Yu K, Fang X. 2023. NGenomeSyn: an easy-to-use and flexible tool for publication-ready visualization of syntenic relationships across multiple genomes. *Bioinformatics* **39**: btad121. doi:10.1093/bioinformatics/btad121
- Huang CY, Ayliffe MA, Timmis JN. 2003. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature* **422**: 72–76. doi:10.1038/nature01435
- Huang CY, Ayliffe MA, Timmis JN. 2004. Simple and complex nuclear loci created by newly transferred chloroplast DNA in tobacco. *Proc Natl Acad Sci* **101**: 9710–9715. doi:10.1073/pnas.0400853101
- Huang CY, Grünheit N, Ahmadinejad N, Timmis JN, Martin W. 2005. Mutational decay and age of chloroplast and mitochondrial genomes transferred recently to angiosperm nuclear chromosomes. *Plant Physiol* **138**: 1723–1733. doi:10.1104/pp.105.060327
- Huang X, Lu G, Zhao Q, Liu X, Han B. 2008. Genome-wide analysis of transposon insertion polymorphisms reveals intraspecific variation in cultivated rice. *Plant Physiol* **148**: 25–40. doi:10.1104/pp.108.121491
- Ju YS. 2016. Intracellular mitochondrial DNA transfers to the nucleus in human cancer cells. *Curr Opin Genet Dev* **38**: 23–30. doi:10.1016/j.gde.2016.02.005
- Ju YS, Tubio JMC, Mifsud W, Fu B, Davies HR, Ramakrishna M, Li Y, Yates L, Gundem G, Tarpey PS, et al. 2015. Frequent somatic transfer of mitochondrial DNA into the nuclear genome of human cancer cells. *Genome Res* **25**: 814–824. doi:10.1101/gr.190470.115
- Kawahara Y, De La Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S, et al. 2013. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* **6**: 4. doi:10.1186/1939-8433-6-4
- Kent WJ. 2002. BLAT—the BLAST-like alignment tool. *Genome Res* **12**: 656–664. doi:10.1101/gr.229202
- Kleinboelting N, Huep G, Appelhagen I, Viehoveer P, Li Y, Weisshaar B. 2015. The structural features of thousands of T-DNA insertion sites are consistent with a double-strand break repair-based insertion mechanism. *Mol Plant* **8**: 1651–1664. doi:10.1016/j.molp.2015.08.011
- Lechardeur D, Sohn K-J, Haardt M, Joshi PB, Monck M, Graham RW, Beatty B, Squire J, O’Brodoovich H, Lukacs GL. 1999. Metabolic instability of plasmid DNA in the cytosol: a potential barrier to gene transfer. *Gene Ther* **6**: 482–497. doi:10.1038/sj.gt.3300867
- Lee JA, Carvalho CMB, Lupski JR. 2007. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* **131**: 1235–1247. doi:10.1016/j.cell.2007.11.037
- Leister D. 2005. Origin, evolution and genetic effects of nuclear insertions of organelle DNA. *Trends Genet* **21**: 655–663. doi:10.1016/j.tig.2005.09.004
- Li W, Freudenberg J, Freudenberg J. 2019. Alignment-free approaches for predicting novel nuclear mitochondrial segments (NUMTs) in the human genome. *Gene* **691**: 141–152. doi:10.1016/j.gene.2018.12.040
- Liang B, Wang N, Li N, Kimball RT, Braun EL. 2018. Comparative genomics reveals a burst of homoplasmy-free numt insertions. *Mol Biol Evol* **35**: 2060–2064. doi:10.1093/molbev/msy112
- Liere K, Weihe A, Börner T. 2011. The transcription machineries of plant mitochondria and chloroplasts: composition, function, and regulation. *J Plant Physiol* **168**: 1345–1360. doi:10.1016/j.jplph.2011.01.005
- Liu S-L, Zhuang Y, Zhang P, Adams KL. 2009. Comparative analysis of structural diversity and sequence evolution in plant mitochondrial genes transferred to the nucleus. *Mol Biol Evol* **26**: 875–891. doi:10.1093/molbev/msp011
- Lukacs GL, Haggie P, Seksek O, Lechardeur D, Freedman N, Verkman AS. 2000. Size-dependent DNA mobility in cytoplasm and nucleus. *J Biol Chem* **275**: 1625–1629. doi:10.1074/jbc.275.3.1625
- Ma X, Fan J, Wu Y, Zhao S, Zheng X, Sun C, Tan L. 2020. Whole-genome *de novo* assemblies reveal extensive structural variations and dynamic organelle-to-nucleus DNA transfers in African and Asian rice. *Plant J* **104**: 596–612. doi:10.1111/tpj.14946
- Marçais G, Delcher AL, Phillippy AM, Costar R, Salzberg SL, Zimin A. 2018. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol* **14**: e1005944. doi:10.1371/journal.pcbi.1005944
- Meyer RS, Choi JY, Sanches M, Plessis A, Flowers JM, Amas J, Dorph K, Barretto A, Gross B, Fuller DQ, et al. 2016. Domestication history and geographical adaptation inferred from a SNP map of African rice. *Nat Genet* **48**: 1083–1088. doi:10.1038/ng.3633
- Min J, Zhao J, Zagalbaum J, Lee J, Takahashi S, Cummings P, Schooley A, Dekker J, Gottesman ME, Rabadan R, et al. 2023. Mechanisms of insertions at a DNA double-strand break. *Mol Cell* **83**: 2434–2448.e7. doi:10.1016/j.molcel.2023.06.016
- Mohanta TK, Mishra AK, Khan A, Hashem A, Abd-Allah EF, Al-Harrasi A. 2020. Gene loss and evolution of the plastome. *Genes (Basel)* **11**: 1133. doi:10.3390/genes11101133
- Morley SA, Nielsen BL. 2017. Plant mitochondrial DNA. *Front Biosci-Landmark* **22**: 1023–1032. doi:10.2741/4531
- Mower JP. 2020. Variation in protein gene and intron content among land plant mitochondrial genomes. *Mitochondrion* **53**: 203–213. doi:10.1016/j.mito.2020.06.002
- Namasivayam S, Sun C, Bah AB, Oberstaller J, Pierre-Louis E, Etheridge RD, Feschotte C, Pritham EJ, Kissinger JC. 2023. Massive invasion of organellar DNA drives nuclear genome evolution in *Toxoplasma*. *Proc Natl Acad Sci* **120**: e2308569120. doi:10.1073/pnas.2308569120
- Noutsos C, Richly E, Leister D. 2005. Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. *Genome Res* **15**: 616–628. doi:10.1101/gr.3788705
- Ntakirutimana F, Tranchant-Dubreuil C, Cubry P, Chougule K, Zhang J, Wing RA, Adam H, Lorieux M, Jouannic S. 2023. Genome-wide association analysis identifies natural allelic variants associated with panicle architecture variation in African rice, *Oryza glaberrima* Steud. *G3 (Bethesda)* **13**: jkad174. doi:10.1093/g3journal/jkad174
- Osia B, Alsulaiman T, Jackson T, Kramara J, Oliveira S, Malkova A. 2021. Cancer cells are highly susceptible to accumulation of templated insertions linked to MMBIR. *Nucleic Acids Res* **49**: 8714–8731. doi:10.1093/nar/gkab685
- Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, et al. 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol* **20**: 275. doi:10.1186/s13059-019-1905-y
- Portugez S, Martin WF, Hazkani-Covo E. 2018. Mosaic mitochondrial-plastid insertions into the nuclear genome show evidence of both non-homologous end joining and homologous recombination. *BMC Evol Biol* **18**: 162. doi:10.1186/s12862-018-1279-x
- Puertas MJ, González-Sánchez M. 2020. Insertions of mitochondrial DNA into the nucleus—effects and role in cell evolution. *Genome* **63**: 365–374. doi:10.1139/gen-2019-0151
- Qin P, Lu H, Du H, Wang H, Chen W, Chen Z, He Q, Ou S, Zhang H, Li X, et al. 2021. Pan-genome analysis of 33 genetically diverse rice accessions

- reveals hidden genomic variations. *Cell* **184**: 3542–3558.e16. doi:10.1016/j.cell.2021.04.046
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842. doi:10.1093/bioinformatics/btq033
- R Core Team. 2023. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>.
- Ricchetti M, Fairhead C, Dujon B. 1999. Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. *Nature* **402**: 96–100. doi:10.1038/47076
- Richly E. 2004. NUMTs in sequenced eukaryotic genomes. *Mol Biol Evol* **21**: 1081–1084. doi:10.1093/molbev/msh110
- Richly E, Leister D. 2004. NUPTs in sequenced eukaryotes and their genomic organization in relation to NUMTs. *Mol Biol Evol* **21**: 1972–1980. doi:10.1093/molbev/msh210
- Rousseau-Gueutin M, Huang X, Higginson E, Ayliffe M, Day A, Timmis JN. 2013. Potential functional replacement of the plastidic acetyl-CoA carboxylase subunit (*accD*) gene by recent transfers to the nucleus in some angiosperm lineages. *Plant Physiol* **161**: 1918–1929. doi:10.1104/pp.113.214528
- Shafer KS, Hanekamp T, White KH, Thorsness PE. 1999. Mechanisms of mitochondrial DNA escape to the nucleus in the yeast *Saccharomyces cerevisiae*. *Curr Genet* **36**: 183–194. doi:10.1007/s002940050489
- Sheppard AE, Timmis JN. 2009. Instability of plastid DNA in the nuclear genome. *PLoS Genet* **5**: e1000323. doi:10.1371/journal.pgen.1000323
- Sheppard AE, Madesis P, Lloyd AH, Day A, Ayliffe MA, Timmis JN. 2011. Introducing an RNA editing requirement into a plastid-localised transgene reduces but does not eliminate functional gene transfer to the nucleus. *Plant Mol Biol* **76**: 299–309. doi:10.1007/s11103-011-9764-2
- Shtolz N, Mishmar D. 2023. The metazoan landscape of mitochondrial DNA gene order and content is shaped by selection and affects mitochondrial transcription. *Commun Biol* **6**: 93. doi:10.1038/s42003-023-04471-4
- Song Y, Wang J. 2023. ggcoverage: an R package to visualize and annotate genome coverage for various NGS data. *BMC Bioinformatics* **24**: 309. doi:10.1186/s12859-023-05438-2
- Song J-M, Lei Y, Shu C-C, Ding Y, Xing F, Liu H, Wang J, Xie W, Zhang J, Chen L-L. 2018. Rice information GateWay: a comprehensive bioinformatics platform for indica rice genomes. *Mol Plant* **11**: 505–507. doi:10.1016/j.molp.2017.10.003
- Song J-M, Xie W-Z, Wang S, Guo Y-X, Koo D-H, Kudrna D, Gong C, Huang Y, Feng J-W, Zhang W, et al. 2021. Two gap-free reference genomes and a global view of the centromere architecture in rice. *Mol Plant* **14**: 1757–1767. doi:10.1016/j.molp.2021.06.018
- Stegemann S, Bock R. 2006. Experimental reconstruction of functional gene transfer from the tobacco plastid genome to the nucleus. *Plant Cell* **18**: 2869–2878. doi:10.1105/tpc.106.046466
- Stegemann S, Hartmann S, Ruf S, Bock R. 2003. High-frequency gene transfer from the chloroplast genome to the nucleus. *Proc Natl Acad Sci* **100**: 8828–8833. doi:10.1073/pnas.1430924100
- Stein JC, Yu Y, Copetti D, Zwickl DJ, Zhang L, Zhang C, Chougule K, Gao D, Iwata A, Goicoechea JL, et al. 2018. Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza*. *Nat Genet* **50**: 285–296. doi:10.1038/s41588-018-0040-0
- Thathapalli Prakash P, Chebotarov D, Zhang J, Kudrna DA, Torres RO, Natividad MA, Quintana MR, Song J, Maldonado CE, Hechanova SL, et al. 2023. *Oryza glumaepatula*: a wild relative to improve drought tolerance in cultivated rice. *Plant Physiol* **193**: 2381–2397. doi:10.1093/plphys/kiad485
- Thorsness PE, Fox TD. 1990. Escape of DNA from mitochondria to the nucleus in *Saccharomyces cerevisiae*. *Nature* **346**: 376–379. doi:10.1038/346376a0
- Uvizl M, Puechmaille SJ, Power S, Pippel M, Carthy S, Haerty W, Myers EW, Teeling EC, Huang Z. 2024. Comparative genome microsynteny illuminates the fast evolution of nuclear mitochondrial segments (NUMTs) in mammals. *Mol Biol Evol* **41**: msad278. doi:10.1093/molbev/msad278
- Vasimuddin Md, Misra S, Li H, Aluru S. 2019. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, Rio de Janeiro, Brazil, pp. 314–324. IEEE, Piscataway, NJ.
- Wang D, Timmis JN. 2013. Cytoplasmic organelle DNA preferentially inserts into open chromatin. *Genome Biol Evol* **5**: 1060–1064. doi:10.1093/gbe/evt070
- Wang D, Lloyd AH, Timmis JN. 2012. Environmental stress increases the entry of cytoplasmic organellar DNA into the nucleus in plants. *Proc Natl Acad Sci* **109**: 2444–2448. doi:10.1073/pnas.1117890109
- Wang D, Gu J, David R, Wang Z, Yang S, Searle IR, Zhu J, Timmis JN. 2018a. Experimental reconstruction of double-stranded break repair-mediated plastid DNA insertion into the tobacco nucleus. *Plant J* **93**: 227–234. doi:10.1111/tj.13769
- Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F, et al. 2018b. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* **557**: 43–49. doi:10.1038/s41586-018-0063-9
- Wei W, Schon KR, Elgar G, Orioli A, Tanguy M, Giess A, Tischkowitz M, Caulfield MJ, Chinnery PF. 2022. Nuclear-embedded mitochondrial DNA sequences in 66,083 human genomes. *Nature* **611**: 105–114. doi:10.1038/s41586-022-05288-7
- Wing RA, Purugganan MD, Zhang Q. 2018. The rice genome revolution: from an ancient grain to green super rice. *Nat Rev Genet* **19**: 505–517. doi:10.1038/s41576-018-0024-z
- Wu Z, Sloan DB, Brown CW, Rosenblueth M, Palmer JD, Ong HC. 2017. Mitochondrial retroprocessing promoted functional transfers of *rpl5* to the nucleus in grasses. *Mol Biol Evol* **34**: 2340–2354. doi:10.1093/molbev/msx170
- Wu Z, Liao X, Zhang X, Tembrock LR, Broz A. 2022. Genomic architectural variation of plant mitochondria—a review of multichromosomal structuring. *J Syst Evol* **60**: 160–168. doi:10.1111/jse.12655
- Xue L, Moreira JD, Smith KK, Fetterman JL. 2023. The mighty NUMT: mitochondrial DNA flexing its code in the nuclear genome. *Biomolecules* **13**: 753. doi:10.3390/biom13050753
- Yoshida T, Furihata HY, Kawabe A. 2017. Analysis of nuclear mitochondrial DNAs and factors affecting patterns of integration in plant species. *Genes Genet Syst* **92**: 27–33. doi:10.1266/ggs.16-00039
- Yu X, Gabriel A. 1999. Patching broken chromosomes with extranuclear cellular DNA. *Mol Cell* **4**: 873–881. doi:10.1016/S1097-2765(00)80397-4
- Yu Z, Chen Y, Zhou Y, Zhang Y, Li M, Ouyang Y, Chebotarov D, Mauleon R, Zhao H, Xie W, et al. 2023. Rice Gene Index: a comprehensive pan-genome database for comparative and functional genomics of Asian rice. *Mol Plant* **16**: 798–801. doi:10.1016/j.molp.2023.03.012
- Yue J, VanBuren R, Liu J, Fang J, Zhang X, Liao Z, Wai CM, Xu X, Chen S, Zhang S, et al. 2022. SunUp and Sunset genomes revealed impact of particle bombardment mediated transformation and domestication history in papaya. *Nat Genet* **54**: 715–724. doi:10.1038/s41588-022-01068-1
- Zelensky AN, Schimmel J, Kool H, Kanaar R, Tijsterman M. 2017. Inactivation of Pol θ and C-NHEJ eliminates off-target integration of exogenous DNA. *Nat Commun* **8**: 66. doi:10.1038/s41467-017-00124-3
- Zhang F, Khajavi M, Connolly AM, Towne CF, Batish SD, Lupski JR. 2009. The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nat Genet* **41**: 849–853. doi:10.1038/ng.399
- Zhang G-J, Dong R, Lan L-N, Li S-F, Gao W-J, Niu H-X. 2020. Nuclear integrants of organellar DNA contribute to genome structure and evolution in plants. *Int J Mol Sci* **21**: 707. doi:10.3390/ijms21030707
- Zhang Y, Tian L, Lu C. 2023. Chloroplast gene expression: recent advances and perspectives. *Plant Commun* **4**: 100611. doi:10.1016/j.xplc.2023.100611
- Zhang X, Chen H, Ni Y, Wu B, Li J, Burzyński A, Liu C. 2024a. Plant mitochondrial genome map (PMGmap): a software tool for the comprehensive visualization of coding, noncoding and genome features of plant mitochondrial genomes. *Mol Ecol Resour* **24**: e13952. doi:10.1111/1755-0998.13952
- Zhang Z, Zhao J, Li J, Yao J, Wang B, Ma Y, Li N, Wang H, Wang T, Liu B, et al. 2024b. Evolutionary trajectory of organelle-derived nuclear DNAs in the *Triticum/Aegilops* complex species. *Plant Physiol* **194**: 918–935. doi:10.1093/plphys/kiad552
- Zhou Y, Chebotarov D, Kudrna D, Llaca V, Lee S, Rajasekar S, Mohammed N, Al-Bader N, Sobel-Sorenson C, Parakkal P, et al. 2020. A platinum standard pan-genome resource that represents the population structure of Asian rice. *Sci Data* **7**: 113. doi:10.1038/s41597-020-0438-2
- Zhou Y, Yu Z, Chebotarov D, Chougule K, Lu Z, Rivera LF, Kathiresan N, Al-Bader N, Mohammed N, Alsanatly A, et al. 2023. Pan-genome inversion index reveals evolutionary insights into the subpopulation structure of Asian rice. *Nat Commun* **14**: 1567. doi:10.1038/s41467-023-37004-y

Received May 27, 2024; accepted in revised form March 27, 2025.