Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Chromosome-resolved genome of Magnaporthe oryzae KJ201 links East Asian rice blast lineage architecture with pathogenicity loci

Hyunjung Chung
National Institute of Crop and Food Science
Yebin Nam
Sunchon National University
Yoeguang Hue
Sunchon National University
Jinyong Kim
Sunchon National University
Byungheon Choi
Sunchon National University
Sunmin An
National Institute of Crop and Food Science
Jaeho Ko
Seoul National University
Yong-Hwan Lee
Seoul National University
Sook-Young Park
Sunchon National University
Ki-Tae Kim
kitaekim@scnu.ac.kr

Sunchon National University

Article

Keywords: Magnaporthe oryzae, comparative genomics, Hi-C scaffolding, secreted effector, structural variation, subtelomere

Posted Date: September 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10888729/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on October 31st, 2026. See the published version at https://doi.org/10.1038/s41598-026-78500-7.

Abstract

Rice blast, caused by *Magnaporthe oryzae*, threatens rice production and is a model for fungal pathogenicity. The Korean rice isolate KJ201 has supported extensive functional, transcriptomic, and epigenomic research, but its fragmented reference genome has limited chromosome-scale interpretation. PacBio HiFi and Hi-C sequencing produced a 44.8 Mb assembly comprising seven chromosome-scale scaffolds and three mini-scaffolds, with an N50 of 6.5 Mb and 98.5% complete BUSCOs. Comparative analysis with 17 public *M. oryzae* genomes placed KJ201 with the East Asian rice isolates P131 and O-135, apart from the 70 − 15/Guy11 lineage. The seven core chromosomes were broadly conserved, whereas localized chromosome-terminal differences and breakpoint patterns differentiated rice subclades and host-associated groups. Gene-family content was largely conserved. Secreted effector candidates showed elevated local repeat density, and 57 ORF-disrupting pathogenicity loci from the KJ201 Agrobacterium-mediated insertional mutant collection were mapped across all seven major chromosomes. This assembly links East Asian lineage architecture, dynamic effector-associated regions, and experimentally derived pathogenicity loci in a native rice blast genome.

INTRODUCTION

Magnaporthe oryzae (anamorph *Pyricularia oryzae*) is the causal agent of rice blast, one of the most destructive diseases of rice worldwide. The fungus has a highly specialized infection cycle involving conidial germination, appressorium-mediated host penetration, invasive growth, and lesion development, making it a major model for fungal morphogenesis, effector biology, and plant-pathogen interactions [1–3]. Because rice blast continues to threaten rice production and food security across diverse agroecosystems, genome-scale resources remain essential for understanding the molecular and evolutionary basis of pathogenicity [4–5].

In *M. oryzae*, pathogenicity and host adaptation are shaped not only by gene content but also by chromosome-scale genome architecture. The first reference genome of strain 70 − 15 provided a foundation for molecular genetics and comparative genomics [6], but short-read and scaffold-level assemblies are often insufficient for resolving repeat-rich compartments, telomeric and subtelomeric regions, accessory sequences, and structural rearrangements. These regions are particularly relevant in blast fungi, where comparative studies have reported gene presence-absence variation, effector turnover, telomere-associated instability, mini-chromosome divergence, horizontally transferred DNA, and host-associated structural variation [7–13]. Chromosome-level assemblies therefore provide more than improved contiguity and they enable direct analysis of how genome organization differs among rice-associated lineages and host pathotypes.

The Korean rice blast strain KJ201 is an important experimental background for functional genomics, infection biology, transcriptomics, and epigenomics. Previous KJ201-based studies examined host interaction phenotypes, transcription factor profiling, infection-stage gene expression, hypoxia-responsive transcriptional regulation, oxidative-stress biology, peroxidase function, developmental DNA methylation, infection-stage transcriptomes, and alternative splicing [14–22]. KJ201 was also used to generate a large-scale *Agrobacterium tumefaciens*-mediated transformation (ATMT) mutant resource, in which 21,070 mutants and 202 new pathogenicity loci were identified through high-throughput phenotypic screening [23]. Despite this extensive functional use, KJ201 has lacked a standardized chromosome-level nuclear genome, limiting the interpretation of mutant phenotypes, transcriptomic and epigenomic signals, effector candidates, and repeat-associated loci in their native chromosomal context. Because KJ201 is a Korean rice isolate, its genome should be interpreted relative to other rice-associated lineages rather than only against the 70 − 15/Guy11 reference lineage [24–25]. A chromosome-level KJ201 reference is therefore needed to connect functional genomics resources with genomic position, synteny, detected telomeric regions, local repeat environment, and lineage- or pathotype-associated chromosome architecture.

We tested two related hypotheses: (i) KJ201 would share a conserved core karyotype with rice-pathotype isolates while retaining localized structural features characteristic of its East Asian lineage, and (ii) chromosome-resolved coordinates would reveal informative genomic contexts for predicted effectors and enable integration of experimentally identified pathogenicity loci. To test these hypotheses, we generated a chromosome-level genome assembly of *M. oryzae* strain KJ201 using PacBio HiFi and Hi-C sequencing and compared it with 17 public chromosome-level *M. oryzae* assemblies representing rice, wheat, finger millet, perennial ryegrass, and tall fescue-associated isolates (Table 1). We evaluated assembly quality, repeat content, phylogenomic placement, chromosome-scale synteny, telomeric sequence recovery, and structural divergence. We then analyzed orthogroup composition, secretome and CAZyme profiles, predicted secreted effector candidates, and ATMT-derived pathogenicity-associated loci. These analyses establish KJ201 as a chromosome-resolved framework for testing how lineage architecture and local genome context shape the organization of rice blast pathogenicity-related genes.

Table 1
Metadata, sequencing platforms, and primary assemblers of *M. oryzae* genome assemblies used in this study.

Isolate	Host	Origin	Sequencing Tech.	Primary Assembler	GenBank Acc.	Reference
KJ201	Rice	South Korea	Illumina	SOAPdenovo v1.05	GCA_000376685.2	[22]
KJ201	Rice	South Korea	PacBio, Hi-C	HiFiasm v0.25.0		This study
P131	Rice	Japan	Illumina, PacBio, Hi-C	Canu v2.1.1	GCA_000292605.2	[26]
O-135	Rice	China	Illumina, ONT	Canu v1.9	GCA_043231925.1	[59]
O-219	Rice	Cote d'Ivoire	Illumina, ONT	Canu v1.9	GCA_043231905.1	[59]
70 – 15	Rice	Lab strain	Illumina, PacBio, Hi-C	Canu v2.2	GCA_049735435.1	[27]
Guy11	Rice	French Guiana	Illumina, ONT	Canu v1.9	GCA_046718735.1	[59]
EA18	Rice	China	Illumina, ONT	NextDenovo v2.4	GCA_021764705.1	[60]
TRG2	Rice	Thailand	PacBio	SPAdes v3.14	GCA_030512215.1	NCBI assembly record
O-137	Rice	China	Illumina, ONT	Canu v1.9	GCA_043231915.1	[59]
95HPH-4	Rice	China	Illumina, PacBio, ONT	Canu v2.2	GCA_050157325.1	[61]
95085	Rice	China	Illumina, ONT	Canu v2.2	GCA_050157855.1	NCBI assembly record
MZ5-1-6	Finger millet	Japan	Illumina, PacBio	HGAP v3	GCA_004346965.1	[10]
Br48	Wheat	Brazil	Illumina, PacBio	HGAP	GCA_036493215.1	[12}
B71	Wheat	Bolivia	Illumina, ONT	Canu v1.9	GCA_004785725.2	[13]
ZM1-2	Wheat	Zambia	Illumina, ONT	Canu v2.2	GCA_026261775.1	NCBI assembly record
T3	Wheat	Brazil	Illumina, ONT	Canu v2.2	GCA_030718865.1	[62]
TF05-1MC7	Tall fescue	USA	Illumina, ONT	Canu v2.1.1	GCA_036172685.1	[32]
LpKY97	Perennial ryegrass	USA	Illumina, ONT	Canu v2.0	GCA_012272995.1	[11]

Unpublished assemblies are identified by their NCBI assembly records.

Results

A chromosome-level KJ201 assembly improves recovery of repeat-rich and chromosome-terminal regions

PacBio HiFi reads and Hi-C data were combined to generate a 44.8 Mb nuclear genome assembly for KJ201. The final assembly consisted of 10 sequences, including seven chromosome-scale scaffolds and three mini-scaffolds (Table 2). The total assembly N50 was 6.51 Mb and the GC content was 51.00%. When the three mini-scaffolds were excluded, the seven major chromosomes comprised 42.86 Mb and retained the same N50, indicating that most nuclear sequence was incorporated into long chromosome-scale molecules. The mini-scaffolds ranged from 44.5 kb to 1.63 Mb, had GC contents of 42.59–51.12% and encoded 8-305 predicted genes. They were retained as separate sequences because their chromosomal placement and orientation could not be resolved unambiguously from the available Hi-C contacts. In contrast, the seven major scaffolds were supported by strong intra-chromosomal interaction blocks with clear boundaries in the Hi-C contact map (Fig. 1A). BUSCO recovered 3,817 of 3,876 sordariomycetes_odb12.2 markers (98.5%), including 3,813 single-copy and four duplicated BUSCOs, no fragmented BUSCOs, and 59 missing BUSCOs (Supplementary Fig. S1). This completeness was comparable to that of the public chromosome-level assemblies analyzed here.

Table 2
Genome assembly and annotation statistics of *M. oryzae* strains.

Isolate	Chr	With Scaf	Chr Size (bp)	Total Size (bp)	Total N50 (bp)	Total GC (%)	Total Repeat content (bp) [%]	Protein-coding Genes	Transcripts
KJ201	-	123	45,096,509	45,096,509	2,318,557	50.91	4,244,937 [10.17]	11,287	12,147
KJ201	7	10	42,858,407	44,771,693	6,512,276	51.00	6,975,054 [15.58]	11,234	12,123
P131	7	7	43,237,743	43,237,743	7,048,921	51.12	6,345,471 [14.68]	11,014	11,839
O-135	7	10	41,960,165	44,468,491	5,950,546	50.97	7,230,394 [16.26]	11,129	11,939
O-219-	7	12	42,648,708	46,767,474	6,382,644	51.01	8,292,709 [17.73]	11,370	12,239
70 – 15	7	7	43,463,497	43,463,497	6,853,756	51.21	7,123,864 [16.39]	10,940	11,731
Guy11	7	12	41,880,509	42,607,926	6,590,161	51.22	6,271,691 [14.72]	10,888	11,675
EA18	7	7	45,746,719	45,746,719	7,493,747	50.95	8,060,043 [17.62]	11,200	12,046
TRG2	7	1,771	36,757,999	38,916,680	5,827,256	51.40	1,706,328 [4.40]	11,081	11,864
O-137	7	8	42,181,937	44,200,050	5,740,312	51.13	6,814,446 [15.42]	11,110	11,951
95HPH-4	7	9	44,955,794	47,777,704	6,738,481	51.00	10,143,049 [21.23]	11,289	12,120
95085	7*	9	45,859,373	45,859,373	6,551,064	50.96	8,120,069 [17.71]	11,185	12,048
MZ5-1-6	7	7	42,703,282	42,703,282	6,129,159	50.09	4,740,374 [11.10]	11,119	11,888
Br48	7	7	42,501,442	42,501,442	6,276,809	50.02	4,203,845 [9.89]	11,155	11,785
B71	7	8	42,908,164	44,811,409	6,668,242	49.85	6,063,840 [13.53]	11,307	11,906
ZM1-2	7	12	41,960,190	44,974,979	6,045,076	49.90	6,183,375 [13.75]	11,355	12,010
T3	7	7	41,682,104	41,682,104	6,306,421	49.97	4,321,590 [10.37]	10,997	11,607
TF05-1MC7	7	11	41,915,541	45,513,680	6,099,538	49.83	6,781,375 [14.90]	11,318	11,985
LpKY97	7	9	41,702,954	45,612,971	6,155,580	49.96	6,386,040 [14.00]	11,460	12,127

*Although nine sequences were initially assigned as chromosomes in strain 95085, BUSCO-based synteny analysis supported seven chromosome-scale groups (Fig. 2).

The new assembly substantially improved the contiguity of the previous Illumina-based KJ201 reference (GCA_000376685.2) [22]. The earlier assembly comprised 123 scaffolds with an N50 of 2.32 Mb, whereas the PacBio HiFi/Hi-C assembly reduced the nuclear genome to seven major chromosomes plus three mini-scaffolds and increased the N50 to 6.51 Mb without materially changing total genome size (Table 2). Whole-genome alignment showed that most sequence from the earlier assembly was retained and assigned to the new chromosomes (Fig. 1B and 1C). Long forward-oriented matches dominated the comparison, while several reverse-oriented blocks and local interruptions indicated corrections to scaffold order or orientation that were unresolved in the short-read assembly. The chromosome-scale assembly therefore provides a more reliable coordinate system while remaining consistent with the overall sequence content of the previous KJ201 genome.

Structural annotation predicted 11,234 protein-coding genes and 12,123 transcripts in KJ201. Repeat annotation identified 6,975,054 bp of repetitive sequence, corresponding to 15.58% of the genome. This was 2.73 Mb more repetitive sequence than detected in the previous Illumina assembly (4.24 Mb), indicating improved recovery of repeat-rich and structurally complex regions (Table 2). Across the chromosome-level *M. oryzae* genomes included in the comparison, repeat content ranged from 9.89% in Br48 to 21.23% in 95HPH-4, excluding the highly fragmented TRG2 assembly, and KJ201 fell within this range (Supplementary Table S1). Most KJ201 repeats were interspersed elements, including unclassified repeats, DNA transposons and retrotransposons. LTR retroelements of the Gypsy/DIRS1 and Ty1/Copia groups formed prominent classified components (Supplementary Fig. S2). Gene density was distributed across all seven major chromosomes, whereas repeat density was more heterogeneous, with localized internal peaks and enrichment toward several chromosome-terminal regions (Fig. 1B).

Canonical TTAGGG/CCCTAA telomeric repeats were detected at 11 terminal regions across eight KJ201 sequences (Fig. 1D). The recovery of telomeric motifs at multiple scaffold ends provides independent support for the long-range continuity of the assembly and confirms that several native chromosome termini were captured. However, telomeric repeats were not detected at both ends of every major chromosome, and the three mini-scaffolds remained unplaced. KJ201 should therefore be considered a high-quality chromosome-level assembly rather than a complete telomere-to-telomere genome.

KJ201 is placed in the East Asian rice C III lineage within a conserved core karyotype

A maximum-likelihood phylogeny based on single-copy orthologs separated rice-pathotype isolates into three well-resolved subclades, designated C I, C II and C III (Fig. 2A). KJ201 clustered with P131 from Japan and O-135 from China in C III, placing the Korean strain within an East Asian rice-associated lineage distinct from the 70 − 15/Guy11 C II lineage. Wheat-pathotype isolates formed a separate clade, while isolates from perennial ryegrass and tall fescue were positioned outside the rice and wheat groups. The phylogenomic placement provided the framework for interpreting chromosome conservation and structural variation among host-associated groups and within rice-pathotype subclades.

BUSCO-marker-based chromosome synteny revealed broad conservation of Chromosomes 1–7 across the analyzed *M. oryzae* assemblies (Fig. 2A). Long forward-oriented syntenic connections predominated, indicating that the seven core chromosomes retain similar large-scale gene order across rice and wheat pathotypes. Collinearity was generally strongest among closely related isolates, and KJ201 showed extensive correspondence with P131 and O-135, consistent with their placement in C III. Nevertheless, several local changes in block orientation, chromosome assignment and continuity were visible within and between rice subclades. These differences were concentrated in restricted regions rather than distributed as chromosome-wide reorganizations, suggesting that structural evolution has occurred within an otherwise conserved core-chromosome framework.

Evolutionary and structural distance summaries supported broad category-level differences among the sampled groups. Most Recent Common Ancestor (MRCA)-to-tip distances differed among rice subclades and the wheat pathotype (Kruskal-Wallis P = 0.021; Fig. 2B). Wheat-pathotype isolates showed the largest distances, whereas C III showed relatively short distances, consistent with KJ201, P131, and O-135 forming a compact lineage. Reciprocal whole-genome alignments also differed in rearrangement burden, expressed as the mean number of synteny breakpoints per 10 Mb of aligned sequence (Kruskal-Wallis P = 1.0×10^{-5}; Fig. 2C). Between-rice-clade comparisons tended to show higher breakpoint burdens than comparisons within a rice clade, whereas the sampled wheat isolates showed the lowest within-group burden and rice-wheat comparisons were intermediate. Because individual isolates occurred in multiple pairwise comparisons and public assemblies differed in continuity, these statistics provide descriptive support for broad structural trends rather than tests based on independent biological replicates. Overall, KJ201 lies within a compact C III lineage embedded in an otherwise conserved core-karyotype framework.

Localized chromosome-terminal differences distinguish C III isolates

One localized C III difference involved KJ201 mini-scaffold 1. A substantial segment showed syntenic correspondence with P131 Chromosome 1, whereas the homologous sequence was represented within a continuous chromosome in P131. Thus, sequence assigned to P131 Chromosome 1 appears to be divided between KJ201 Chromosome 1 and mini-scaffold 1. The mini-scaffold therefore contains core-chromosome-homologous sequence and may represent an unresolved chromosome-terminal segment or a candidate chromosome-end difference. Because repeat-rich terminal sequence can reduce the power of Hi-C scaffolding, assembly ambiguity and biological structural variation cannot yet be separated at this locus.

A second localized difference involved the 3' terminal region of P131 Chromosome 2, which aligned in reverse orientation to Chromosome 7 in both KJ201 and O-135. The homologous block therefore differed in chromosomal assignment and orientation among the C III assemblies. The shared Chromosome 7 placement in KJ201 and O-135 and the Chromosome 2 placement in P131 identify a candidate chromosome-terminal configuration difference, but the ancestral state and exact breakpoint mechanism cannot be determined from three assemblies alone. Validation will require breakpoint-spanning long reads or independent physical mapping.

Pairwise MUMmer dotplots independently supported the chromosome-scale collinearity of KJ201 with representative rice-pathotype assemblies (Supplementary Fig. S3). Most major chromosomes produced long, continuous positive-slope diagonals, confirming conservation of chromosome order and orientation. The correspondence between KJ201 mini-scaffold 1 and P131 Chromosome 1 appeared as a discrete alignment block separated from the principal Chromosome 1 diagonal, supporting the presence of Chromosome 1-homologous sequence on the mini-scaffold. Likewise, the P131 Chromosome 2-terminal/KJ201 Chromosome 7 relationship produced a localized reverse-oriented block rather than a chromosome-wide anti-diagonal. Numerous shorter off-diagonal matches were also present and are likely to include repetitive elements, duplicated sequence and assembly-specific placements. Their dispersed distribution contrasted with the larger coherent reverse-oriented blocks interpreted as candidate structural rearrangements. No evidence was found for extensive whole-chromosome inversions or widespread chromosome reshuffling in KJ201.

Telociraptor detected between one and 16 telomeric repeat regions per public assembly (Fig. 2A). KJ201 contained 11 detected regions across eight sequences, compared with 12 in P131, 14 in O-135 and 12 in 70 – 15 (Supplementary Table S2). All motifs belonged to the canonical TTAGGG/CCCTAA family and were located predominantly at scaffold ends. Because telomere detection depends strongly on assembly continuity and recovery of terminal repeats, these values were interpreted as profiles of chromosome-end sequence recovery rather than direct estimates of biological telomere or chromosome number.

A conserved orthogroup backbone contains limited lineage-specific gene content

Orthology analysis across the *M. oryzae* isolates and the outgroup *M. grisea* NI907 identified 14,454 orthogroups distributed among 1,326 distinct presence-absence intersections (Supplementary Table S3 and S4). The largest component consisted of 8,922 orthogroups present in every analyzed strain, indicating extensive conservation of the *Magnaporthe* gene-family repertoire (Fig. 3A). A further 670 orthogroups were present in all *M. oryzae* isolates but absent from NI907 and were designated the *M. oryzae*-specific core. In contrast, strictly host- or lineage-restricted sets were much smaller, 41 orthogroups were specific to the rice pathotype, 92 to non-rice isolates, and only two to the wheat pathotype. Within rice isolates, no orthogroup was specific to C I, one was specific to C II and four were specific to C III. Nineteen strain-specific intersections contained 2,138 orthogroups in total, including 25 singleton orthogroups assigned to KJ201. These counts show that most gene families are shared broadly, whereas strict pathotype, clade and strain specificity is confined to a relatively small accessory fraction.

Exploratory GO enrichment of the 670 *M. oryzae*-specific core orthogroups highlighted redox and cofactor metabolism, lipid metabolism, secondary metabolism and carbohydrate-active functions (Fig. 3B). Representative terms included flavin adenine dinucleotide, heme, iron ion and NADP binding, monooxygenase and O-methyltransferase activities, triacylglycerol lipase and polygalacturonase activities, and secondary-metabolite biosynthetic processes. These functions are consistent with the broad metabolic and enzymatic capabilities required during environmental growth and host colonization. Rice-pathotype-specific orthogroups showed candidate enrichment for fatty acid-CoA ligase and sterol reductase activities, phospholipid translocation, xylan degradation, autophagy and TORC1-related signaling (Fig. 3C). Together, these terms suggest that a limited set of rice-associated genes may influence membrane remodeling, nutrient utilization and regulatory processes rather than forming a single specialized virulence pathway.

KJ201 singleton orthogroups produced a smaller and less stable enrichment profile, including candidate terms related to DNA-templated transcription, phosphorelay sensor kinase activity, phosphorelay signaling and septum-associated processes (Fig. 3D). Because both the singleton and rice-subclade-specific sets contained few genes, their enrichments were interpreted as exploratory signals based on nominal statistics. Direct inspection of the subclade-specific proteins identified predicted transferase-, membrane- and proteolysis-related proteins among the four C III orthogroups represented by KJ201, while one protein lacked detectable functional annotation (Table 3). The single C II-specific orthogroup represented by 70 – 15 encoded a short hypothetical protein without assigned GO terms. Thus, differentiation among rice lineages occurs primarily against a conserved gene-family backbone, with a small number of candidate lineage-specific genes available for functional testing. Because strict lineage-specific orthogroups were few, we next asked whether pathogenicity-related differentiation was reflected more strongly in genomic context than in gene-family expansion.

Table 3
Functional annotation of representative Rice C II- and Rice C III-specific genes.

Clade Type	Representative Strain	Protein ID	Description	Length (aa)	GO IDs	GO Names
C_II-specific	70 – 15	g9170.t1	Hypothetical protein MCOR02_007688	66	-	-
C_III-specific	KJ201	g284.t2	Uncharacterized protein MGG_02269	86	GO:0016740	Transferase activity
C_III-specific	KJ201	g9093.t1	-	55	-	-
C_III-specific	KJ201	g10479.t2	Hypothetical protein OOU_Y34scaffold00301g29	387	GO:0016020	Membrane
C_III-specific	KJ201	g11022.t1	Hypothetical protein OOU_Y34scaffold00956g9, partial	336	GO:0006508 GO:0004252 GO:0016020	Proteolysis Serine-type endopeptidase activity Membrane

Secreted effector candidates show elevated repeat density in their genomic neighborhoods

Secretome, CAZyme and effector predictions were compared across 18 *M. oryzae* genomes to determine whether KJ201 displayed unusual expansion or reduction of pathogenicity-associated protein classes. KJ201 encoded 1,603 SignalP-predicted secreted proteins, a value within the range of the other strains (Fig. 4A). The overall distribution of glycoside hydrolases, glycosyltransferases, polysaccharide lyases, carbohydrate esterases, auxiliary activity enzymes and carbohydrate-binding modules was also broadly conserved. However, secretome filtering markedly altered the apparent contribution of some CAZyme classes. KJ201 contained 122 glycosyltransferases, but only five carried predicted signal peptides, consistent with most fungal glycosyltransferases acting intracellularly and only a small subset participating directly in extracellular host interaction.

EffectorP3 applied to the complete KJ201 proteome predicted 248 apoplastic, 1,951 cytoplasmic and 85 mixed apoplastic/cytoplasmic candidates (Fig. 4A). Because genome-wide application produced a particularly large cytoplasmic candidate set, subsequent positional analyses were restricted to proteins supported by both SignalP and EffectorP. This conservative intersection contained 385 secreted effector candidates, including 186 apoplastic, 157 cytoplasmic and 42 mixed candidates. The filtering step does not verify secretion or host translocation, but it reduces the influence of intracellular proteins lacking an N-terminal signal peptide and provides a more focused set for chromosome-context comparisons.

Most secreted effector candidates had moderate 5' and 3' flanking intergenic distances, although a subset occurred in regions with extended flanking spaces and therefore occupied relatively gene-sparse neighborhoods (Fig. 4B). Chromosome-wide mapping showed that apoplastic, cytoplasmic and mixed candidates were distributed across all seven major chromosomes and the mini-scaffolds rather than confined to a single genomic compartment (Fig. 4C). Repeat density, however, was strongly heterogeneous. Several candidate effectors occurred near repeat-enriched chromosome ends or localized internal repeat peaks, suggesting that part of the effector repertoire is embedded in regions with greater potential for recombination, transposon activity or structural change.

Descriptive repeat-density profiles supported this positional pattern. Across windows extending 50 kb on either side of each locus, all three secreted effector classes showed higher mean repeat density than the genome-wide background set (Fig. 4D). Cytoplasmic and mixed candidates also showed broader distributions toward highly repetitive neighborhoods in the per-locus comparison (Fig. 4E). Thus, KJ201 does not exhibit a major increase in total secretome or CAZyme content, but its conservative secreted effector candidates occur in genomic regions with elevated local repeat density. These descriptive patterns identify candidate dynamic genomic environments; they do not establish that the predicted proteins are secreted during infection, translocated into rice cells, required for virulence, or statistically enriched relative to independent biological replicates.

The chromosome-level reference anchors legacy pathogenicity loci

To connect the new KJ201 annotation with the widely used 70 – 15 gene nomenclature, KJ201 proteins were matched to MGG proteins by bidirectional DIAMOND searches. Candidate MGG assignments were obtained for 11,536 of 12,123 KJ201 protein models, corresponding to 10,307 unique MGG identifiers (Supplementary Table S5). Reciprocal best-hit analysis identified 10,279 high-confidence KJ201-MGG pairs, representing 84.8% of all KJ201 protein models and 89.1% of KJ201 proteins with a candidate MGG match. These pairs generally showed near-complete sequence identity and query and subject coverage. A further 1,257 proteins retained only single-direction candidate mappings, while 587 proteins lacked detectable MGG assignments. The reciprocal set therefore provides a conservative cross-reference, with the remaining mappings retained for cases affected by gene-model differences, duplication or lineage-specific sequence divergence.

Of the 202 pathogenicity-associated loci reported in the original KJ201 ATMT screen, 57 loci with T-DNA insertions within annotated ORFs were converted from MGG identifiers to KJ201 coordinates and projected onto the chromosome-level assembly [23]. These ORF-disrupting loci were distributed across all seven major chromosomes rather than concentrated in a single chromosome or terminal region, and none mapped to the mini-scaffolds (Fig. 5A). Some insertions were associated only with reduced pathogenicity, whereas many also caused defects in vegetative growth, pigmentation, conidiation, conidial germination, conidial morphology, or appressorium formation. The chromosome-level map therefore shows that loss of pathogenicity in the original screen frequently accompanied disruption of broader developmental or cellular functions.

To determine whether the chromosome-mapped pathogenicity loci converged on shared functional modules, the mapped genes were queried against the MAGNET association network. Twenty-four genes formed a functional-association network when edges with scores greater than three were retained (Fig. 5B). The network connected genes associated with pathogenicity-only phenotypes to genes showing pleiotropic pathogenicity and developmental defects. Highly connected nodes, including MGG_06700 (TBC domain-containing protein), linked multiple phenotype classes, suggesting that some loci may participate in shared regulatory or cellular modules rather than acting as isolated virulence determinants. Genes affecting conidiation and appressorium formation were interspersed with other pathogenicity genes, consistent with the dependence of successful infection on coordinated asexual development, infection structure formation, and invasive growth.

GO annotations of the 24 MAGNET-connected genes were diverse and did not converge on a single pathway (Fig. 5C). Frequently represented terms included nucleus, cytoplasm, membrane, mitochondrion, vacuolar membrane, zinc ion binding, DNA binding, ATP binding, GTPase activator activity, and proteolysis. Additional annotations involved transcriptional control, DNA repair, actin cytoskeleton organization, vesicle-mediated transport, lipid metabolism, TOR signaling, and RNA splicing. The word-frequency summary is descriptive and does not constitute an enrichment analysis. Together with the MAGNET network, it provides a hypothesis-generating framework for prioritizing KJ201 loci, not evidence of direct physical interactions or a single pathogenicity pathway.

Discussion

Our results support both hypotheses. KJ201 shares a broadly conserved core karyotype with rice-pathotype isolates while retaining localized chromosome-terminal features within the East Asian C III lineage, and its chromosome-resolved coordinates provide a framework for integrating predicted secreted effectors and experimentally derived pathogenicity loci. PacBio HiFi and Hi-C data reduced the previous 123-scaffold assembly to seven chromosome-scale scaffolds and three mini-scaffolds, increased the N50 from 2.3 to 6.5 Mb, and recovered approximately 2.7 Mb of additional repetitive sequence while maintaining 98.5% BUSCO completeness. The value of this improvement extends beyond continuity. The original 70 − 15 genome established *M. oryzae* as a model for fungal plant pathogenesis and revealed a large secreted-protein repertoire and abundant transposable elements [6], but fragmented assemblies make it difficult to determine whether candidate genes lie in stable chromosome interiors, subtelomeres, or rearranged repeat-rich regions. Recent telomere-to-telomere and near-complete assemblies of rice blast strains have further shown that chromosome ends and previously unresolved repeats can materially alter annotation and structural interpretation [26–27]. KJ201 is not yet telomere-to-telomere, but it now supplies a chromosome-resolved coordinate system for integrating mutant insertions, expression profiles, DNA methylation, alternative splicing, repeats, and candidate effector loci generated in this widely used experimental background.

The phylogenomic analysis also demonstrates why a KJ201-specific reference is needed. KJ201 grouped with P131 and O-135 in the East Asian rice-associated C III lineage rather than with 70 − 15 and Guy11 in C II. Rice blast populations contain multiple endemic and pandemic lineages with distinct demographic histories [25], and genome comparisons across host groups have documented recurrent gene gain and loss, effector turnover and chromosomal change [10, 24]. Accordingly, interpreting KJ201 only against 70 − 15 risks conflating lineage-level differences with KJ201-specific features. In the present analysis, the seven core chromosomes remained broadly collinear across rice isolates, but reciprocal breakpoint burden increased in many between-clade comparisons. This combination of conserved chromosome identity and localized structural divergence is consistent with evolution acting through regional inversions, translocations and chromosome-end changes rather than wholesale reorganization of the core karyotype. The particularly low burden within the sampled wheat group should not be taken as evidence that the wheat pathotype is intrinsically less dynamic, because the result may also reflect narrower sampling, recent common ancestry and differences in assembly quality.

The C III comparisons highlight two chromosome-terminal features of particular relevance to KJ201. First, mini-scaffold 1 contains sequence homologous to P131 Chromosome 1, indicating that at least part of this scaffold is derived from a core chromosome region rather than representing wholly accessory sequence. Second, a terminal segment located on P131 Chromosome 2 aligns in reverse orientation to Chromosome 7 in both KJ201 and O-135. Similar movement of effector-containing segments, exchange between core and dispensable chromosomes and repeat-associated loss of synteny have been reported in blast fungi [28–31]. Transposon-mediated telomere instability and the rapid remodeling of mini-chromosomes provide plausible mechanisms for chromosome-end variation [11, 13, 32]. Nevertheless, the present data do not identify the causal mechanism. Repeat-rich terminal regions can be difficult to scaffold, and a shared Chromosome 7 placement in KJ201 and O-135 does not by itself establish whether P131 is derived or whether an unobserved ancestral configuration existed. The two regions should therefore be treated as well-supported candidate structural differences requiring breakpoint-spanning read validation or physical chromosome analysis.

Gene-family comparisons suggest that lineage differentiation occurs mainly through modification of a shared genomic repertoire rather than wholesale acquisition of novel gene families. Most orthogroups were conserved across *M. oryzae* and *M. grisea* NI907, whereas strict pathotype-, rice-subclade- and KJ201-specific sets were small. Host adaptation may therefore depend on combinations of allelic divergence, gene copy-number change, regulatory evolution, chromosomal position and presence-absence variation at a limited number of loci. This interpretation agrees with studies showing

that blast-fungus host specialization is associated with dynamic gain and loss of genes linked to transposable elements and with diversification of candidate effector repertoires among pathotypes [8–9]. The functional enrichments detected here, including redox, lipid and secondary metabolism, carbohydrate utilization, autophagy and TOR-associated processes, are biologically plausible in the context of fungal growth and infection, but they should not be overinterpreted. Several accessory gene sets were small, and enrichment was used as an exploratory summary rather than evidence that a complete pathway is uniquely present or expanded in a lineage. The scarcity of strict lineage-specific orthogroups therefore shifts the main comparative question from large-scale gene-family acquisition to allelic, regulatory, positional, and repeat-associated variation within a shared repertoire.

The secretome and effector analyses further illustrate the importance of combining gene prediction with genomic context. KJ201 did not show a marked expansion of its total secretome or CAZyme complement, supporting broad conservation of the extracellular enzymatic repertoire among the strains analyzed. EffectorP alone produced a large number of genome-wide candidates, particularly in the cytoplasmic class, so the positional analysis was restricted to proteins also predicted to carry signal peptides. Even within this conservative set, secreted effector candidates showed elevated local repeat density relative to the descriptive genomic background. This result is consistent with the recognized role of transposable elements in effector gain, loss and translocation in *M. oryzae* and with recent population-scale evidence that individual AVR loci occupy distinct TE-defined evolutionary niches [32–33]. Chromatin state may add a further regulatory layer, because H3K27-associated changes have been linked to altered expression of in planta-induced genes [34]. However, repeat proximity is a descriptive genomic association rather than proof of rapid evolution or virulence function. Infection-stage expression, secretion and translocation assays, targeted gene disruption and host-genotype testing will be required to distinguish bona fide effectors from computational candidates.

A major practical outcome of the assembly is the chromosome-level reinterpretation of the KJ201 ATMT resource. The original screen of more than 21,000 insertion mutants identified over 200 loci associated with pathogenicity and demonstrated that virulence defects frequently coincide with abnormalities in conidiation, germination, appressorium formation, and vegetative growth [23]. Mapping 57 ORF-disrupting loci to KJ201 chromosomes now enables these phenotypes to be compared with local repeat density, structural variation, expression, and epigenomic data in the same genetic background. The MAGNET network connected genes involved in transcription, signaling, membrane trafficking, cytoskeletal organization, metabolism, and organelle function, reinforcing the view that pathogenicity is an emergent cellular phenotype rather than the output of a single virulence pathway. Important limitations remain: three mini-scaffolds are unplaced, the genome is not fully telomere-to-telomere, pairwise breakpoint estimates are affected by assembly quality and non-independent comparisons, repeat-density comparisons are descriptive, and orthogroup, effector, GO, and network analyses are computational. Non-ORF ATMT insertions were also excluded from the chromosome plot. Population-level chromosome assemblies from Korean and other East Asian isolates, combined with breakpoint-spanning reads and targeted functional validation, will be needed to resolve the terminal variants and test candidate genes. Within these constraints, the KJ201 reference bridges comparative and functional genomics by linking East Asian lineage architecture, repeat-associated effector neighborhoods, and legacy pathogenicity phenotypes within a single native genome.

Methods

Fungal culture and DNA preparation

M. oryzae strain KJ201 was maintained on V8 juice agar containing 8% V8 juice and 1.5% agar, adjusted to pH 6.9. Mycelial plugs from a 7-day-old culture were transferred to fresh 9 cm plates containing 25 mL medium and incubated at 25°C for 7 days. Mycelial biomass was shipped to Phyzen (Seongnam, Republic of Korea) for high-molecular-weight DNA extraction, quality control and sequencing-library preparation. Genomic DNA was isolated using the Wizard HMW

DNA Extraction Kit (Promega, Madison, WI, USA) according to the manufacturer protocol. DNA concentration was measured with PicoGreen (Invitrogen, Waltham, MA, USA), and fragment size distribution and integrity were evaluated using a Femto Pulse system (Agilent, Santa Clara, CA, USA).

PacBio and Hi-C sequencing

High-molecular-weight DNA was sheared with a Megaruptor 3 system (Diagenode, Denville, NJ, USA), purified and size-selected using AMPure PB beads (Pacific Biosciences, Menlo Park, CA, USA). SMRTbell libraries with an approximate insert-size range of 7–12 kb were prepared using the SMRTbell Prep Kit 3.0 and sequenced on the PacBio Revio platform with the Revio SPRQ polymerase kit. The resulting circular-consensus HiFi reads were used for de novo assembly. A Hi-C library was prepared independently with the Proximo Plant Hi-C Kit v4.5 (Phase Genomics, Seattle, WA, USA) and sequenced as paired-end reads on an Illumina NovaSeq 6000. The Hi-C data were used to order and orient PacBio contigs and to identify candidate scaffold misjoins.

Genome assembly and chromosome-level scaffolding

PacBio HiFi reads were assembled de novo with Hifiasm v0.25.0-r726 using default parameters [35]. Putative mitochondrial contigs were identified by alignment to the Guy11 mitochondrial genome (NC_066489.1) with minimap2 v2.28-r1209 and removed from the nuclear assembly [36]. The remaining contigs were processed with purge_dups v1.2.6 to remove haplotypic duplication and redundant sequence [37]. Assembly size, sequence number, N50, maximum sequence length and GC content were calculated with SeqKit v2.13.0 [38].

Hi-C reads were quality-filtered with Fastp v0.23.2 and aligned to the curated PacBio contigs with BWA-MEM v0.7.17-r1188 [39–40]. PCR and optical duplicates were marked and removed using Picard MarkDuplicates v3.4.0. SAMtools was then used to exclude unmapped, secondary and supplementary alignments, and only alignments with mapping quality of at least 10 were retained [41]. These filtered contact pairs were used as the input for chromosome-scale scaffolding.

Scaffolding was performed with YaHS v1.2.2 [42]. Hi-C contact maps were generated from the YaHS output using Juicer v1.2.2 and Juicer Tools v2.20.00 and inspected in Juicebox v2.17.00 [43]. Scaffold order and orientation were manually curated, and candidate misjoins were corrected according to the interaction map. The final nuclear assembly was evaluated with SeqKit and BUSCO v6.1.0 using the sordariomycetes_odb12.2 lineage dataset [44]. Sequences that could not be placed confidently by Hi-C evidence were retained as mini-scaffolds rather than assigned arbitrarily to a chromosome.

Repeat, gene, and functional annotation

For comparative annotation, the KJ201 assembly and 17 public *M. oryzae* assemblies downloaded from NCBI were processed through the same repeat-masking and gene-prediction workflow (Table 1). A genome-specific repeat library was generated for each assembly with RepeatModeler v2.0.5 using the LTR structural-discovery module [45]. RepeatMasker v4.1.6 was then run with the corresponding library to classify and mask interspersed repeats [46]. This common workflow was used to reduce differences caused by heterogeneous annotations in the public records.

Protein-coding genes were predicted with BRAKER v3.0.7 in fungal mode [47]. The repeat-masked assemblies were analyzed using intrinsic gene prediction together with protein-homology evidence from 858,797 Pyriculariaceae protein sequences from NCBI. AUGUSTUS predictions and evidence-supported models were integrated to produce the final structural annotation [48]. Predicted proteins were searched against the NCBI non-redundant protein database (25 September 2025), and Gene Ontology terms were assigned with the mapping and annotation workflow in OmicsBox v4.0.54.

Signal peptides were predicted with SignalP v6.0g, candidate apoplastic and cytoplasmic effectors with EffectorP v3.0, and CAZymes with dbCAN v3.0.7 [49–51]. EffectorP classes were summarized for the complete proteome. For chromosome-position, intergenic-distance and repeat-association analyses, secreted effector candidates were defined conservatively as proteins predicted both to contain a signal peptide and to belong to an EffectorP effector class. These proteins were treated as computational candidates and not as experimentally validated effectors.

Phylogenomic inference and evolutionary distance analysis

Orthogroups were inferred from the predicted protein sets with OrthoFinder v2.5.4 [52]. Single-copy orthologs were aligned individually with MAFFT v7.490 [53], and the alignments were concatenated for maximum-likelihood phylogenomic inference with IQ-TREE v3.0.1 [54]. ModelFinder was used to select the best-fitting substitution model, and support was assessed with 1,000 ultrafast bootstrap and 1,000 SH-aLRT replicates. The tree was rooted with M. grisea NI907. For each rice subclade and the wheat-pathotype clade, the branch-length distance from the inferred MRCA to each terminal isolate was calculated in substitutions per site. Differences among groups were summarized with the Kruskal-Wallis test.

Orthogroup presence-absence classification and exploratory GO analysis

Orthogroup presence was scored when at least one predicted protein from an isolate was assigned to an orthogroup. Orthogroups present in all analyzed *M. oryzae* isolates and *M. grisea* NI907 were classified as the universal core, whereas orthogroups present in all *M. oryzae* isolates and absent from NI907 were classified as the *M. oryzae*-specific core. Rice-pathotype-, wheat-pathotype-, rice-subclade-, and strain-specific sets were defined strictly by presence in every member of the focal group and absence from all nonfocal isolates. Strain-specific singleton sets were restricted to a single isolate. Presence-absence intersections were summarized from the OrthoFinder output and reported in Supplementary Tables S3 and S4.

GO terms assigned through OmicsBox were summarized for the *M. oryzae*-specific core, rice-pathotype-specific orthogroups, and KJ201-specific singletons. GO-term over-representation outputs were ranked by gene ratio and nominal P value, and the leading terms were visualized as descriptive summaries. Because the rice-subclade and singleton sets contained few genes, these analyses were treated as exploratory; no pathway was considered uniquely enriched or expanded on the basis of nominal P values alone.

Chromosome-scale synteny and telomeric repeat detection

Chromosome-scale synteny was analyzed with ChromSyn v1.6.1 using conserved BUSCO markers from the sordariomycetes_odb12.2 dataset [44, 55]. Marker order and chromosome assignment were compared among assemblies, and forward- and reverse-oriented syntenic relationships were plotted in phylogenomic order. Telomeric regions were detected with Telociraptor v0.11.0 followed by manual sequence-level review. Both orientations of the canonical TTAGGG/CCCTAA motif were searched, and detected regions were overlaid on the synteny panel. Telomere counts were used as indicators of terminal sequence recovery rather than direct biological chromosome counts.

Pairwise whole-genome dotplots were generated with nucmer v3.1 from MUMmer [56]. KJ201 was aligned to representative rice-pathotype genomes using maxmatch option, a minimum cluster length of 500 bp and a minimum exact-match length of 100 bp. Delta files were filtered to retain alignments of at least 5 kb and at least 90% nucleotide identity. Coordinates were extracted with show-coords, and plots were produced with chromosome order fixed according to the input FASTA files. Continuous positive-slope diagonals were interpreted as collinear sequence, whereas coherent negative-slope blocks provided evidence for reverse-oriented alignments. Short dispersed off-diagonal matches were interpreted cautiously because they may represent repeats or duplicated sequence.

Pairwise synteny breakpoint analysis

Chromosome structural divergence was quantified from reciprocal whole-genome alignments. Each genome pair was aligned in both directions with minimap2 v2.28-r1209 using the asm10 preset and with secondary alignments disabled [36]. Alignment blocks shorter than 10 kb, below 80% identity or with mapping quality (MAPQ) below 30 were removed. Synteny breakpoints were counted when adjacent retained blocks changed query chromosome or contig, changed orientation, showed a large discontinuity in query coordinates or were separated by a long target gap. Breakpoint burden was normalized as the number of breakpoints per 10 Mb of aligned sequence. The final value for a genome pair was the mean of the two reciprocal directions. Comparisons were grouped as within rice clade, between rice clades, within wheat and rice-wheat, and category differences were assessed with the Kruskal-Wallis test. Because pairs sharing an isolate are not independent, P values were interpreted descriptively.

Intergenic-distance and repeat-density analyses

For positional analysis, 5' and 3' flanking distances were measured from each secreted effector candidate to the nearest annotated feature on the same sequence. Repeat coverage was calculated from the RepeatMasker intervals. Non-overlapping 10-kb windows were used for chromosome-scale repeat-density maps. Gene-centered profiles summarized repeat coverage around each candidate coordinate, and mean repeat density within ± 50 kb was calculated for the per-locus comparison. The genome-wide annotated set served as the descriptive background. Because loci on the same chromosomes share genomic neighborhoods and do not represent independent biological replicates, effector-background repeat-density differences were presented descriptively rather than subjected to an inferential significance test.

KJ201-to-70-15 protein identifier matching

KJ201 predicted proteins were matched to the 70 – 15 MG8 MGG protein set with DIAMOND BLASTP v0.9.30.131 [57]. Searches were performed in both directions using an E-value threshold of 10 and a maximum of 25 target hits per query. Hits were ranked by bit score, E-value, alignment length and sequence identity. Pairs that were the highest-ranking match in both search directions were classified as reciprocal best hits and used as high-confidence KJ201-MGG mappings. Single-direction best hits were retained as candidate assignments, while many-to-one or one-to-many relationships were flagged as potentially ambiguous. The resulting table was used to translate legacy MGG identifiers to the new KJ201 gene models.

ATMT locus projection and MAGNET network analysis

Pathogenicity-associated loci and phenotype categories were obtained from the KJ201 ATMT screen of Jeon et al. [23]. MGG identifiers were converted to KJ201 coordinates using the protein-matching table. The chromosome map included the 57 loci for which the reported T-DNA insertion occurred within an annotated open reading frame. Insertions classified as intergenic, adjacent or boundary-associated were excluded from the plotted set. Phenotype classes included growth rate, pigmentation, conidiation, conidial germination, appressorium formation, conidial morphology and pathogenicity. Mapped genes were queried against the MAGNET Magnaporthe functional-association network, and edges with scores greater than three were visualized [58]. Network links were interpreted as predicted functional associations and not as evidence of direct physical or genetic interaction.

Statistical analysis

Kruskal–Wallis tests were used to assess differences among groups in MRCA-to-tip distance and reciprocal synteny breakpoint burden. For breakpoint comparisons, genome pairs sharing an isolate were not considered independent biological replicates; therefore, the resulting P values were interpreted descriptively. GO enrichment analyses were

performed using Fisher's exact test with a nominal significance threshold of $P < 0.05$. Because no terms remained significant after false discovery rate (FDR) correction, the GO results were interpreted as exploratory.

Declarations

Ethics declarations

Conflict of interest

The authors declare no conflicts of interest.

Funding

This research was carried out with the support of "Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ017573012026)", Rural Development Administration, Republic of Korea.

Author Contribution

H.C.: Validation, Investigation, Writing - Original Draft, Funding acquisition. Y.N.: Methodology, Formal analysis, Investigation, Writing - Original Draft. Y.H.: Formal analysis, Data curation, Visualization. J.K.: Data curation, Visualization. B.C.: Software, Formal analysis. S.A.: Investigation. J.K.: Investigation. Y.-H.L.: Writing - Review and Editing. S.-Y.P.: Resources, Writing - Review and Editing. K.-T.K.: Conceptualization, Validation, Investigation, Data curation, Writing - Original Draft, Writing - Review and Editing, Visualization, Supervision, Project administration.

Acknowledgement

We thank the research community for generating and publicly sharing genomic resources for M. oryzae and related blast fungus isolates. These publicly available datasets provided an important foundation for the comparative genomic and phylogenomic analyses performed in this study.

Data Availability

The chromosome-level genome assembly of M. oryzae strain KJ201 has been deposited in NCBI GenBank under accession number JCAJWW000000000. The raw sequencing data are available from the NCBI Sequence Read Archive under BioProject PRJNA1465931 and BioSample SAMN59837073, with PacBio HiFi and Hi-C reads deposited under SRR38604269 and SRR38604268, respectively. Structural annotation files, repeat annotation files, gene and protein sequences, orthogroup tables, and other supporting datasets are available through Figshare (https://doi.org/10.6084/m9.figshare.32934119).

References

1. Talbot, N. J. On the trail of a cereal killer: exploring the biology of *Magnaporthe grisea*. *Annu. Rev. Microbiol.* **57**, 177–202 (2003).
2. Wilson, R. A. & Talbot, N. J. Under pressure: investigating the biology of plant infection by *Magnaporthe oryzae*. *Nat. Rev. Microbiol.* **7**, 185–195 (2009).

3. Fernandez, J. & Orth, K. Rise of a cereal killer: the biology of *Magnaporthe oryzae* biotrophic growth. *Trends Microbiol.* **26**, 582–597 (2018).
4. Baudin, M. et al. *Pyricularia oryzae*: Lab star and field scourge. *Mol. Plant. Pathol.* **25**, e13449 (2024).
5. Hue, Y. et al. Changes in the occurrence patterns of rice fungal diseases due to climate change. *Res. Plant. Dis.* **31**, 17–29 (2025).
6. Dean, R. A. et al. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434**, 980–986 (2005).
7. Gowda, M. et al. Genome analysis of rice-blast fungus *Magnaporthe oryzae* field isolates from southern India. *Genom Data*. **5**, 284–291 (2015).
8. Yoshida, K. et al. Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. *BMC Genom.* **17**, 370 (2016).
9. Kim, K. T. et al. Evolution of the genes encoding effector candidates within multiple pathotypes of *Magnaporthe oryzae*. Front. Microbiol. 10, 2575 (2019).
10. Gómez Luciano, L. B. et al. Blast fungal genomes show frequent chromosomal changes, gene gains and losses, and effector gene turnover. *Mol. Biol. Evol.* **36**, 1148–1161 (2019).
11. Rahnama, M. et al. Transposon-mediated telomere destabilization: a driver of genome evolution in the blast fungus. *Nucleic Acids Res.* **48**, 7197–7217 (2020).
12. Kobayashi, N. et al. Horizontally transferred DNA in the genome of the fungus *Pyricularia oryzae* is associated with repressive histone modifications. *Mol. Biol. Evol.* **40**, msad186 (2023).
13. Liu, S. et al. Rapid mini-chromosome divergence among fungal isolates causing wheat blast outbreaks in Bangladesh and Zambia. *New. Phytol*. **241**, 1266–1276 (2024).
14. Park, J. Y., Jin, J., Lee, Y. W., Kang, S. & Lee, Y. H. Rice blast fungus (*Magnaporthe oryzae*) infects Arabidopsis via a mechanism distinct from that required for the infection of rice. *Plant. Physiol.* **149**, 474–486 (2009).
15. Kim, S., Park, J., Park, S. Y., Mitchell, T. K. & Lee, Y. H. Identification and analysis of in planta expressed genes of *Magnaporthe oryzae*. *BMC Genom.* **11**, 104 (2010).
16. Park, S. Y. et al. Global expression profiling of transcription factor genes provides new insights into pathogenicity and stress responses in the rice blast fungus. *PLoS Pathog*. **9**, e1003350 (2013).
17. Kim, S. et al. Homeobox transcription factors are required for conidiation and appressorium development in the rice blast fungus *Magnaporthe oryzae*. *PLoS Genet.* **5**, e1000757 (2009).
18. Choi, J. et al. Genome-wide analysis of hypoxia-responsive genes in the rice blast fungus, *Magnaporthe oryzae*. PLoS One 10, e0134939 (2015).
19. Mir, A. A. et al. Systematic characterization of the peroxidase gene family provides new insights into fungal pathogenicity in *Magnaporthe oryzae*. *Sci. Rep.* **5**, 11831 (2015).
20. Jeon, J. et al. Genome-wide profiling of DNA methylation provides insights into epigenetic regulation of fungal development in a plant pathogenic fungus, *Magnaporthe oryzae*. *Sci. Rep.* **5**, 8567 (2015).
21. Jeon, J. et al. Transcriptome profiling of the rice blast fungus *Magnaporthe oryzae* and its host Oryza sativa during infection. Mol. *Plant-Microbe Interact.* **33**, 141–144 (2020).
22. Jeon, J. et al. Alternative splicing diversifies the transcriptome and proteome of the rice blast fungus during host infection. *RNA Biol.* **19**, 373–386 (2022).
23. Jeon, J. et al. Genome-wide functional analysis of pathogenicity genes in the rice blast fungus. *Nat. Genet.* **39**, 561–565 (2007).
24. Chiapello, H. et al. Deciphering genome content and evolutionary relationships of isolates from the fungus *Magnaporthe oryzae* attacking different host plants. *Genome Biol. Evol.* **7**, 2896–2912 (2015).

25. Gladieux, P. et al. Coexistence of multiple endemic and pandemic lineages of the rice blast pathogen. *mBio* **9**, e01806–e01817 (2018).
26. Li, Z. et al. First telomere-to-telomere gapless assembly of the rice blast fungus *Pyricularia oryzae*. *Sci. Data*. **11**, 380 (2024).
27. Cheng, H. Y., Jiang, L. P., Fei, Y., Lu, F. & Ma, S. An annotated near-complete sequence assembly of the *Magnaporthe oryzae* 70 – 15 reference genome. *Sci. Data*. **12**, 758 (2025).
28. Thon, M. R. et al. The role of transposable element clusters in genome evolution and loss of synteny in the rice blast fungus *Magnaporthe oryzae*. *Genome Biol.* **7**, R16 (2006).
29. Chuma, I. et al. Multiple translocation of the AVR-Pita effector gene among chromosomes of the rice blast fungus *Magnaporthe oryzae* and related species. *PLoS Pathog*. **7**, e1002147 (2011).
30. Peng, Z. et al. Effector gene reshuffling involves dispensable mini-chromosomes in the wheat blast fungus. *PLoS Genet.* **15**, e1008272 (2019).
31. Langner, T. et al. Genomic rearrangements generate hypervariable mini-chromosomes in host-specific isolates of the blast fungus. *PLoS Genet.* **17**, e1009386 (2021).
32. Lin, G. et al. Highly dynamic supernumerary mini-chromosomes in a *Magnaporthe oryzae* strain contributes to cellular variance of genomic content. *Fungal Genet. Biol.* **180**, 104020 (2025).
33. Sampaio, A. M. & Croll, D. Transposable elements create distinct genomic niches for effector evolution among *Magnaporthe oryzae* lineages. *BMC Biol.* **23**, 282 (2025).
34. Zhang, W., Huang, J. & Cook, D. E. Histone modification dynamics at H3K27 are associated with altered transcription of in planta induced genes in *Magnaporthe oryzae*. *PLoS Genet.* **17**, e1009376 (2021).
35. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods*. **18**, 170–175 (2021).
36. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
37. Guan, D. et al. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898 (2020).
38. Shen, W., Sipos, B. & Zhao, L. SeqKit2: A Swiss army knife for sequence and alignment processing. *iMeta* **3**, e191 (2024).
39. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
40. Li, H. & arXiv Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. (2013). https://arxiv.org/abs/1303.3997
41. Danecek, P. et al. Twelve years of SAMtools and BCFtools. GigaScience 10, giab008 (2021).
42. Zhou, C., McCarthy, S. A. & Durbin, R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39**, btac808 (2023).
43. Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell. Syst.* **3**, 99–101 (2016).
44. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
45. Flynn, J. M. et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. U.S.A. 117, 9451–9457 (2020).
46. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinformatics 25, 4.10.1–4.10.14 (2009).

47. Gabriel, L. et al. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res.* **34**, 769–777 (2024).
48. Hoff, K. J. & Stanke, M. Predicting genes in single genomes with AUGUSTUS. *Curr. Protoc. Bioinf.* **65**, e57 (2019).
49. Sperschneider, J. & Dodds, P. N. EffectorP 3.0: prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. *Mol. Plant-Microbe Interact.* **35**, 146–156 (2022).
50. Teufel, F. et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* **40**, 1023–1025 (2022).
51. Zheng, J. et al. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Res.* **51**, W115–W121 (2023).
52. Emms, D. M., Liu, Y., Belcher, L., Holmes, J. & Kelly, S. OrthoFinder: improved phylogenetic orthology inference with enhanced accuracy and scalability. *Nat. Methods*. **23**, 1327–1333 (2026).
53. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
54. Wong, T. K. F. et al. IQ-TREE 3: phylogenomic inference software using complex evolutionary models. *Mol. Biol. Evol.* **43**, msag117 (2026).
55. Edwards, R. J., Dong, C., Park, R. F., Tobias, P. A. & bioRxiv A phased chromosome-level genome and full mitochondrial sequence for the dikaryotic myrtle rust pathogen, Austropuccinia psidii. (2022). https://doi.org/10.1101/2022.04.22.489119
56. Kurtz, S. et al. Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).
57. Buchfink, B., Reuter, K. & Drost, H. G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods*. **18**, 366–368 (2021).
58. Kim, H. et al. MagNet: computational methods for constructing high-confidence protein-protein interaction networks in *Magnaporthe oryzae*. bioRxiv (2026). https://doi.org/10.64898/2026.05.11.724438
59. Rowe, D. et al. Natural genomic variation in rice blast genomes is associated with specific heterochromatin modifications. (2023). https://doi.org/10.1101/2023.08.30.555587

Figures

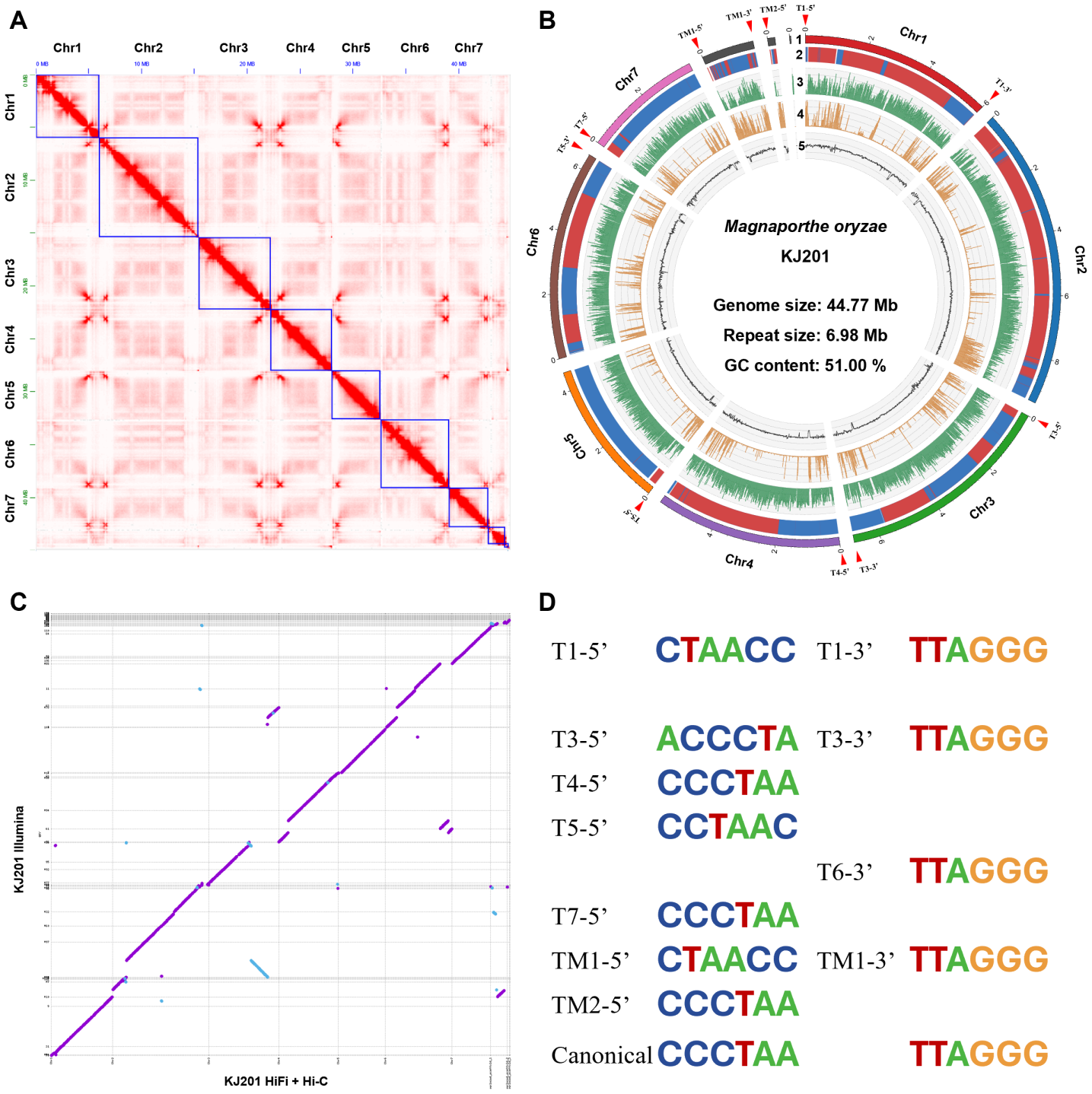


Figure 1

Chromosome-level genome assembly and genomic landscape of *M. oryzae* strain KJ201. A, Hi-C contact map of the final KJ201 assembly. Blue boxes indicate the major chromosome-scale interaction blocks. B, Circos representation of the KJ201 chromosome-level genome assembly. Track 1 shows the final KJ201 genome assembly, including seven major chromosomes and three mini-scaffolds. Track 2 shows alignment of the previous Illumina-based KJ201 assembly to the new chromosome-level assembly; blue and red blocks indicate forward and reverse alignments, respectively. Track 3 shows gene density calculated in 10-kb windows. Track 4 shows repeat density calculated in 10-kb windows. Track 5 shows GC content. Red triangles indicate detected telomeric repeat motifs at chromosome ends. C,

MUMmer dotplot comparing the chromosome-level KJ201 assembly with the previous Illumina-based KJ201 assembly. D, Telomeric repeat sequences detected in the KJ201 assembly.

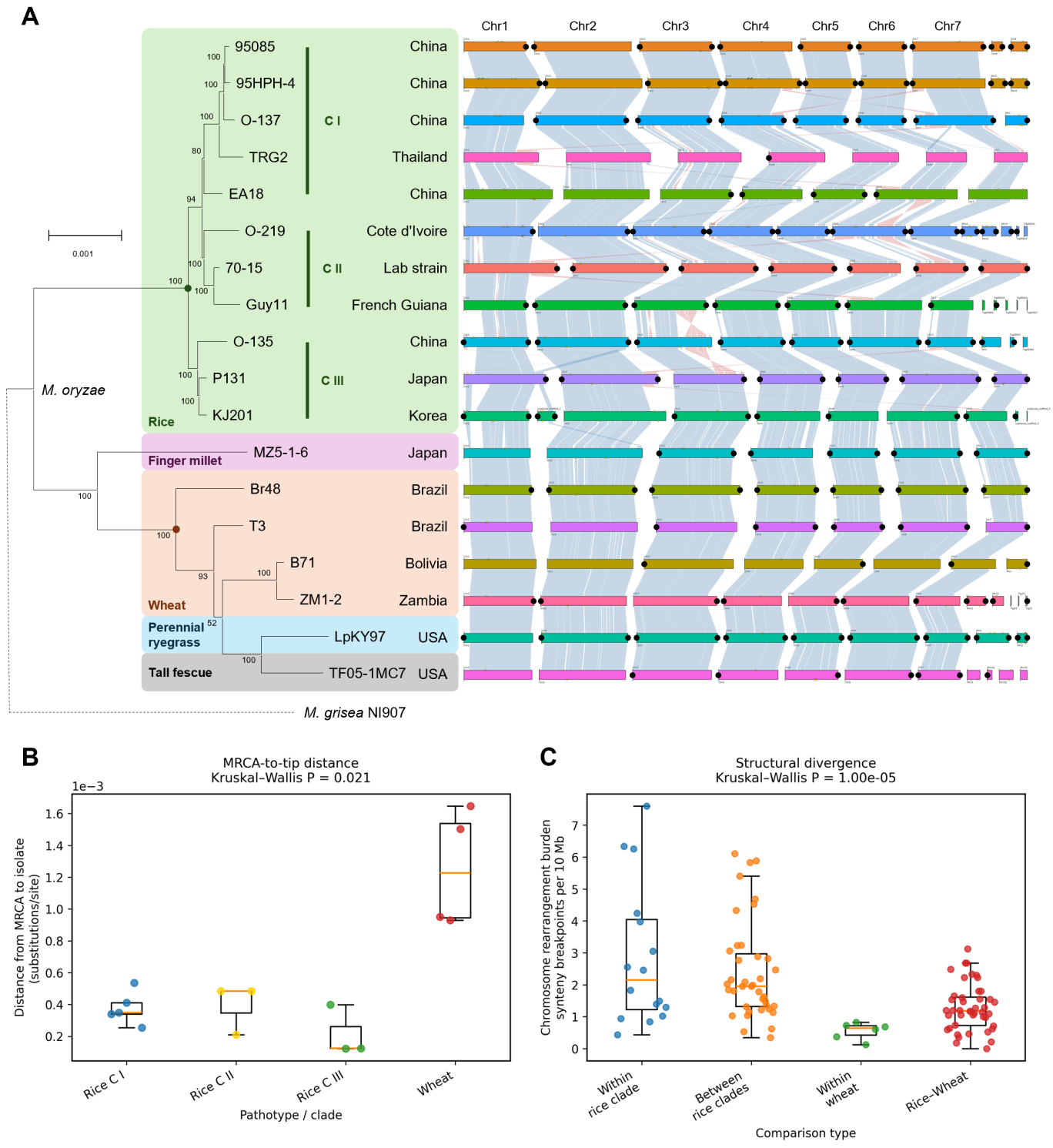


Figure 2

Phylogenomic relationships, chromosome-scale synteny, and structural divergence among *M. oryzae*strains. A, Phylogenomic tree and chromosome-scale synteny comparison of representative *M. oryzae* strains. Rice-pathotype isolates were divided into three subclades, designated C I, C II, and C III. Wheat-pathotype isolates formed a separate

clade, whereas perennial ryegrass and tall fescue isolates were placed outside the rice and wheat groups. Host groups are indicated by background colors. Chromosome-scale synteny across Chr1-Chr7 was inferred using the ChromSyn BUSCO-based approach and visualized according to phylogenomic tree order. Blue and red syntenic connections indicate forward and reverse relationships, respectively. Black dots indicate telomeric regions detected by Telociraptor. Branch nodes marked with dots indicate the inferred most recent common ancestors used for MRCA-to-tip distance calculations. The dashed representation of *M. grisea* NI907 is not drawn to scale. For TRG2, only chromosome-level scaffolds are shown, and unanchored contigs are omitted. B, MRCA-to-tip distance for rice clades and wheat-pathotype isolates. Distances were calculated as substitutions per site from the inferred most recent common ancestor to each terminal isolate. C, Pairwise chromosome structural divergence among comparison types. Rearrangement burden was calculated as the reciprocal pairwise mean number of synteny breakpoints per 10 Mb of aligned sequence. Comparisons were grouped as within rice clade, between rice clades, within wheat, and rice-wheat.

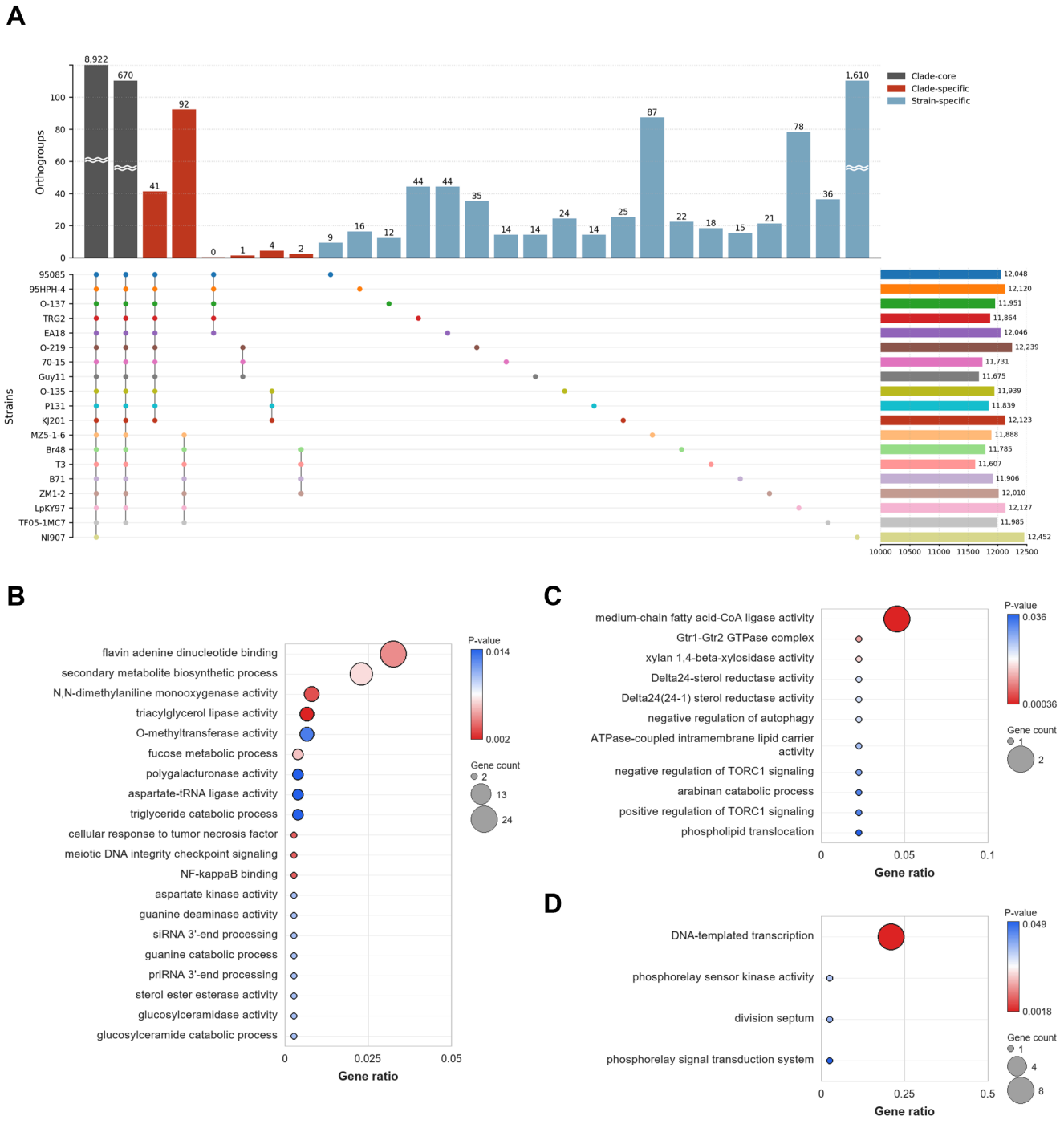


Figure 3

Conserved and lineage-restricted orthogroups in *M. oryzae*. A, UpSet plot of shared and specific orthogroups among *M. oryzae* isolates and the outgroup *M. grisea* NI907. Dark gray, red, and blue bars represent core, clade/pathotype-specific, and strain-specific orthogroups, respectively. Horizontal bars indicate the number of predicted proteins in each strain. B, Exploratory GO-term analysis of *M. oryzae*-specific core orthogroups. The top 20 terms are shown. C, Exploratory GO-term analysis of rice pathotype-specific orthogroups. D, Exploratory GO-term analysis of KJ201-specific singletons. Dot size indicates the number of genes assigned to each term, and color indicates the nominal P value. Small target sets and nominal statistics require cautious interpretation.

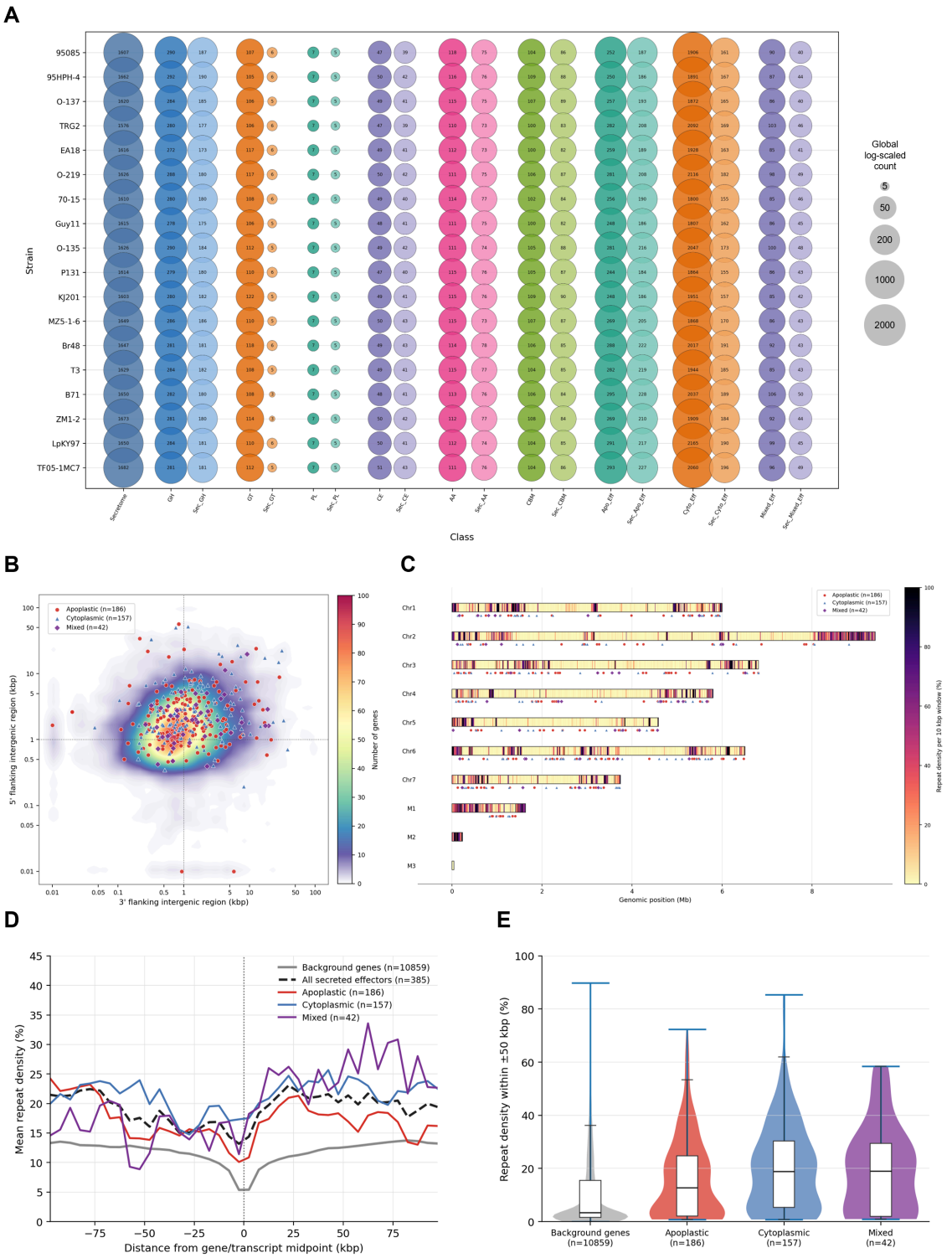


Figure 4

Comparative secretome, CAZyme, and secreted effector candidate profiles in *M. oryzae* strain KJ201. A, Bubble plot showing predicted secretome size, CAZyme class counts, EffectorP3-predicted effector class counts, and their secretome-overlapped subsets across 18 *M. oryzae* strains. Each row represents a strain, and each column represents either the total number of proteins in a class or the subset overlapping with the predicted secretome. Sec indicates secreted proteins. CAZyme classes include GH, glycoside hydrolases; GT, glycosyltransferases; PL, polysaccharide lyases; CE, carbohydrate esterases; AA, auxiliary activity enzymes; and CBM, carbohydrate-binding modules. Effector

classes are Apo, apoplastic; Cyto, cytoplasmic; and Mixed, apoplastic/cytoplasmic. Bubble size indicates log-scaled protein count, and values inside bubbles indicate the number of proteins. Downstream analyses in panels B-E were restricted to candidates supported by both SignalP and EffectorP3. B, Distribution of KJ201 secreted effector candidates according to 5' and 3' flanking intergenic region lengths. The background density represents genome-wide annotated loci. C, Chromosome-wide distribution of repeat density and secreted effector candidate positions in KJ201. Repeat density was calculated in 10-kb windows and displayed along each chromosome and mini-scaffold. Colored markers indicate secreted apoplastic, cytoplasmic, and mixed candidates. D, Mean repeat density around locus midpoints for the genomic background, all secreted effector candidates, and each candidate subclass. The dashed vertical line indicates the locus midpoint. E, Distribution of mean repeat density within ±50 kb of background loci and secreted effector candidates. Violin and box plots show the descriptive distributions.

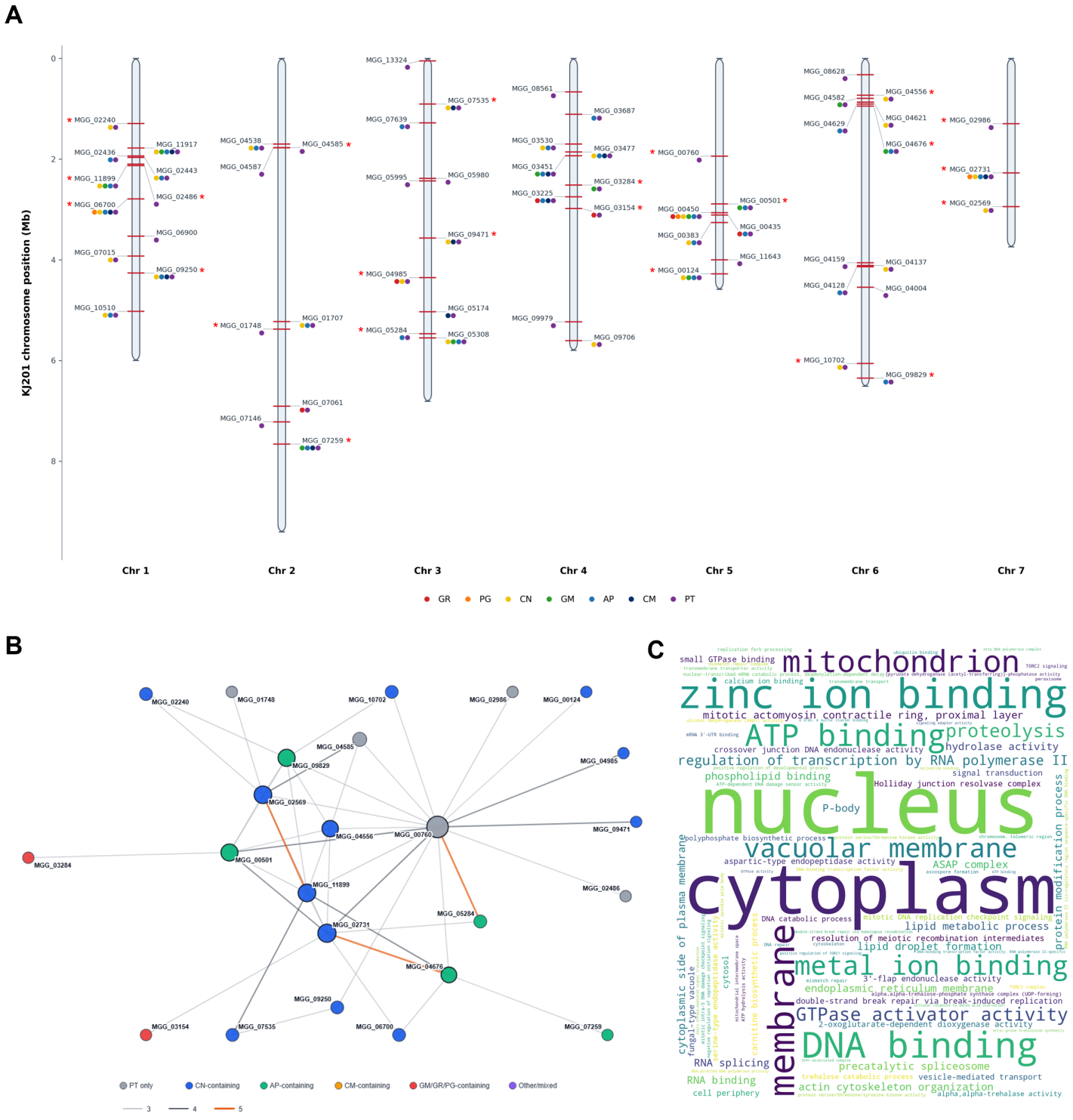


Figure 5

Chromosomal distribution and functional network of pathogenicity-associated ATMT loci in KJ201. A, Chromosomal distribution of 57 pathogenicity-associated genes from the ATMT mutant dataset projected onto the KJ201 chromosome-level genome assembly. Only loci with T-DNA insertions within ORFs are shown. Red ticks indicate mapped positions, colored dots indicate mutant phenotype categories, and asterisks indicate genes included in the MAGNET network shown in panel B. Phenotype abbreviations are GR, growth rate; PG, pigmentation; CN, conidiation; GM, conidial germination; AP, appressorium formation; CM, conidial morphology; and PT, pathogenicity. B, MAGNET

network of 24 pathogenicity-associated genes with network scores greater than 3. Nodes represent genes, node colors indicate phenotype classes, and edges indicate predicted functional associations rather than direct physical interactions. C, Word cloud summarizing GO terms assigned to the 24 MAGNET-connected genes. Word size reflects term frequency; this panel is a descriptive frequency summary rather than a GO enrichment analysis.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigures.docx
- SupplementaryTables.xlsx