

RESEARCH

Open Access



Genome-wide expansion and reorganization during grass evolution: from 30 Mb chromosomes in rice and *Brachypodium* to 550 Mb in *Avena*

Qing Liu^{1,2,3*}, Lyuhan Ye^{1,4}, Mingzhi Li⁵, Ziwei Wang⁶, Gui Xiong^{1,4}, Yushi Ye^{1,2}, Tiejiao Tu^{1,2,3}, Trude Schwarzacher^{1,7} and John Seymour (Pat) Heslop-Harrison^{1,7*}

Abstract

Background The BOP (Bambusoideae, Oryzoideae, and Pooideae) clade of the Poaceae has a common ancestor, with similarities to the genomes of rice, *Oryza sativa* ($2n = 24$; genome size 389 Mb) and *Brachypodium distachyon* ($2n = 10$; 271 Mb). We exploit chromosome-scale genome assemblies to show the nature of genomic expansion, structural variation, and chromosomal rearrangements from rice and *Brachypodium*, to diploids in the tribe Aveneae (e.g., *Avena longiglumis*, $2n = 2x = 14$; 3,961 Mb assembled to 3,850 Mb in chromosomes).

Results Most of the *Avena* chromosome arms show relatively uniform expansion over the 10-fold to 15-fold genome-size increase. Apart from non-coding sequence diversification and accumulation around the centromeres, blocks of genes are not interspersed with blocks of repeats, even in subterminal regions. As in the tribe Triticeae, blocks of conserved synteny are seen between the analyzed species with chromosome fusion, fission, and nesting (insertion) events showing deep evolutionary conservation of chromosome structure during genomic expansion. Unexpectedly, the terminal gene-rich chromosomal segments (representing about 50 Mb) show translocations between chromosomes during speciation, with homogenization of genome-specific repetitive elements within the tribe Aveneae. Newly-formed intergenomic translocations of similar extent are found in the hexaploid *A. sativa*.

Conclusions The study provides insight into evolutionary mechanisms and speciation in the BOP clade, which is valuable for measurement of biodiversity, development of a clade-wide pangenome, and exploitation of genomic diversity through breeding programs in Poaceae.

Keywords Ancestral karyotype, *Avena*, Chromosomal rearrangements, Genomic expansion, Oat, Retrotransposons, Structural variation, Translocations

*Correspondence:

Qing Liu

liuqing@scib.ac.cn

John Seymour (Pat) Heslop-Harrison

phh4@le.ac.uk

¹Key Laboratory of Plant Resources Conservation and Sustainable Utilization, Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China

²South China National Botanical Garden, Guangzhou 510650, China

³Center for Conservation Biology, Core Botanical Gardens, Chinese Academy of Sciences, Guangzhou 510650, China

⁴University of Chinese Academy of Sciences, Beijing 100049, China

⁵Bio&Data Biotechnologies Co. Ltd, Guangzhou 510663, China

⁶Henry Fok School of Biology and Agriculture, Shaoguan University, Shaoguan 512005, China

⁷Department of Genetics and Genome Biology, Institute for Environmental Futures, University of Leicester, Leicester LE1 7RH, UK



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Genomic studies have shed light on the nature and processes of gene evolution, with variation in DNA and RNA sequence data enabling development of robust phylogenies [1]. Multiple polyploidy or whole-genome duplication (WGD) events have played a major part in plant speciation and genome evolution [2, 3] with the separation of Poales from other monocotyledonous orders around 60–110 million years ago (Mya) [4]. The ρ WGD event occurred 50–70 Mya, at the end of the Cretaceous period [5–7] and marked separation of the BOP (Bambusoideae, Oryzoideae, and Pooideae) clade, which includes rice, oats, and wheat, from other grass lineages [8–11]. In contrast to other angiosperm families such as Brassicaceae [12, 13], further WGD events in the BOP clade occurred much more recently [14], and many polyploids in both Triticeae (wheat) and Aveneae (oats) arose in the last few million years [15]. Based on palaeogenomic research, Murat et al. (2010) proposed an ancestral grass karyotype (AGK) based on 5 to 7 chromosomes, with the post- ρ WGD karyotype having 12 chromosome pairs, similar to extant *Oryza sativa* (rice, $2n=2x=24$) and with derived numbers down to $n=5$ (*Brachypodium distachyon*, Brachypodium, $2n=2x=10$) [16]. Rice has preserved the AGK [6], and also like Brachypodium, has a small genome size with neither the transposon activities nor the repeat accumulation observed elsewhere in the grasses. Genome size shows very substantial variation in the BOP clade, from 271 Mb in Brachypodium and 389 Mb in rice [17, 18], to more than 4,000 Mb in many diploid Triticeae and a similar size in Aveneae (both $x=7$) [19, 20]. Genome variation and chromosome reorganization have been shown to be important in plant breeding [2–4]. Further work is needed to understand the interplay between repetitive DNA proliferation, insertion/retention bias in the BOP clade, and how to harness this biology to enhance traits of agronomic importance.

For larger genomes, genome-scale analyses of evolution have been hampered by sequence mis-assembly in scaffolds (and linkage breakage), fragmented assemblies, as well as collapse of similar reads of repetitive motifs from thousands of copies to a small number during assembly [21, 22]. Long-read sequencing technologies (e.g., from Oxford Nanopore Technologies, ONT, or PacBio) combined with genome scaffolding methods (e.g., high-throughput chromatin conformation capture, Hi-C) now enable inclusion of the majority of repetitive DNA sequences in contiguous genome assemblies that reach chromosome-scale [23–26].

Here we aimed to characterize the nature of genome-size expansion within the grass BOP clade, from the small genomes of rice and Brachypodium, represented as being close to the AGK, to the magnitude larger genomes of diploid *Avena* species, integrating conserved synteny and

chromosome-block evolution. In addition, we addressed whether there are genomic hot-spots of integration of repetitive elements, and whether genes remain in compact blocks during genomic expansion. Our results offer genomic insights into the evolutionary history of the Pooideae grasses and the basis for developing the grass pangenome.

Materials and methods

Genome sequences

A chromosome-scale genome assembly of *Avena longiglumis* (ALO; PI 657387; US Department of Agriculture at Beltsville, <https://www.ars-grin.gov/>), originally collected in Morocco) are deposited into the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) under accession number PRJNA956334. The chromosome-scale genome assemblies of *A. atlantica* (AAT) and *A. eriantha* (AER) were downloaded from <https://genomeevolution.org/coge/GenomeInfo.pl?gid=53337> and <https://genomeevolution.org/coge/GenomeInfo.pl?gid=53381>, respectively [20]. A chromosome-scale genome assembly of *Brachypodium distachyon* (BDI) was downloaded from NCBI SRA (PRJNA32607) [17]. A chromosome-scale genome assembly of *Oryza sativa* (OSA) was downloaded from https://phytozome-next.jgi.doe.gov/info/Osativa_v7_0 [18].

Genome annotation

Repeat analysis

De novo repeat prediction was carried out across all the reference genomes. We found that the use of published annotations from different versions of software resulted in unsatisfactory comparisons of identifications and abundances of repetitive elements. For the ALO assembly, repeat prediction was carried out by EDTA v.1.7.0 (Extensive *de-novo* TE Annotator [27], which was composed of eight software modules. The LTRharvest (LTRharvest, PRID:SCR_018970) [28], LTR_FINDER_parallel (LTR_FINDER, PRID:SCR_015247) [29], LTR_retriever [27], Generic Repeat Finder [30] and TIR-Learner [31] modules were included to identify TIR transposons. The HelitronScanner v.1.0 [32] module was used to identify *Helitron* transposons. The RepeatModeler v.2.0.2a [33] module was used to identify TEs (such as *LINES*). Finally, the RepeatMasker v.4.1.1 (RepeatMasker, RRID:SCR_012954) [34] module was used to annotate fragmented TEs based on homology to structurally annotated TEs. In addition, the TESorter v.1.1.4 [35] module was used to identify TE-related genes (Additional file 2: Table S8). The final set of repetitive sequences in the ALO assembly was obtained by integrating *ab initio*-predicted TEs and those identified by homology through RepeatMasker (Additional file 2: Table S8A). Intact LTR-RTs

were identified using LTR_retriever [28]. For comparison, the same repeat analysis protocol was applied to other two grass genomes, *B. distachyon* [17] and *O. sativa* [18] (Additional file 2: Table S8B), in the context of genome size. All LTR-RT families were clustered based on their LTR sequences.

Centromere locations in ALO were identified by the following genomic features: (1) high abundances of repeat sequences on chromosome dotplots (Additional file 1: Figure S3); (2) discontinuities in the Hi-C contact map (Figure S4 in Liu et al. [36]); (3) locations of barley (*Hordeum vulgare*) Gypsy LTR *Cereba* (KM948610) [37] sequence used to identify centromeres in wheat (*Triticum aestivum*, TAE) [38] [the *Cereba* sequence was aligned to the ALO assembly using BLASTN and the centromere cores were identified using Geneious Prime v.2021.1.1 (<https://www.geneious.com/>; Additional file 2: Table S5B)]; (4) SynVisio [39] visualization of gaps and conserved regions between the ALO and OSA assemblies (Fig. 1, Additional file 2: Table S5C); and regions of low gene density along each ALO chromosome. Centromeric cores were defined by overlapping high-abundance repeat

regions on ALO chromosome dotplots and regions of low gene density on ALO chromosomes.

Gene family identification and phylogenetic tree reconstruction

To evaluate evolution and divergence of the genome assembly, protein-coding gene sequences from six species, ALO, AAT [20], AER [20], *A. strigosa* (AST) [40], BDI [17], and OSA [18], were downloaded from Phytozome v.13 [41] and NCBI (<https://www.ncbi.nlm.nih.gov/>) for comparative analyses (Additional file 2: Tables S3 and S4). When one gene had multiple transcripts, only the longest transcript in the coding region was kept for further analysis. Paralogs and orthologs were clustered with OrthoFinder v.2.3.14 [42], using standard parameters, with Diamond v.0.9.24 [43]. Single-copy of orthologous genes were extracted from OrthoFinder [42].

Orthogroup analysis

We used standard methods to discover syntenic blocks of genes [44, 45]. Protein sequences within and between genomes were searched against one another to detect putative homologous genes (E value $< 1e-5$) using

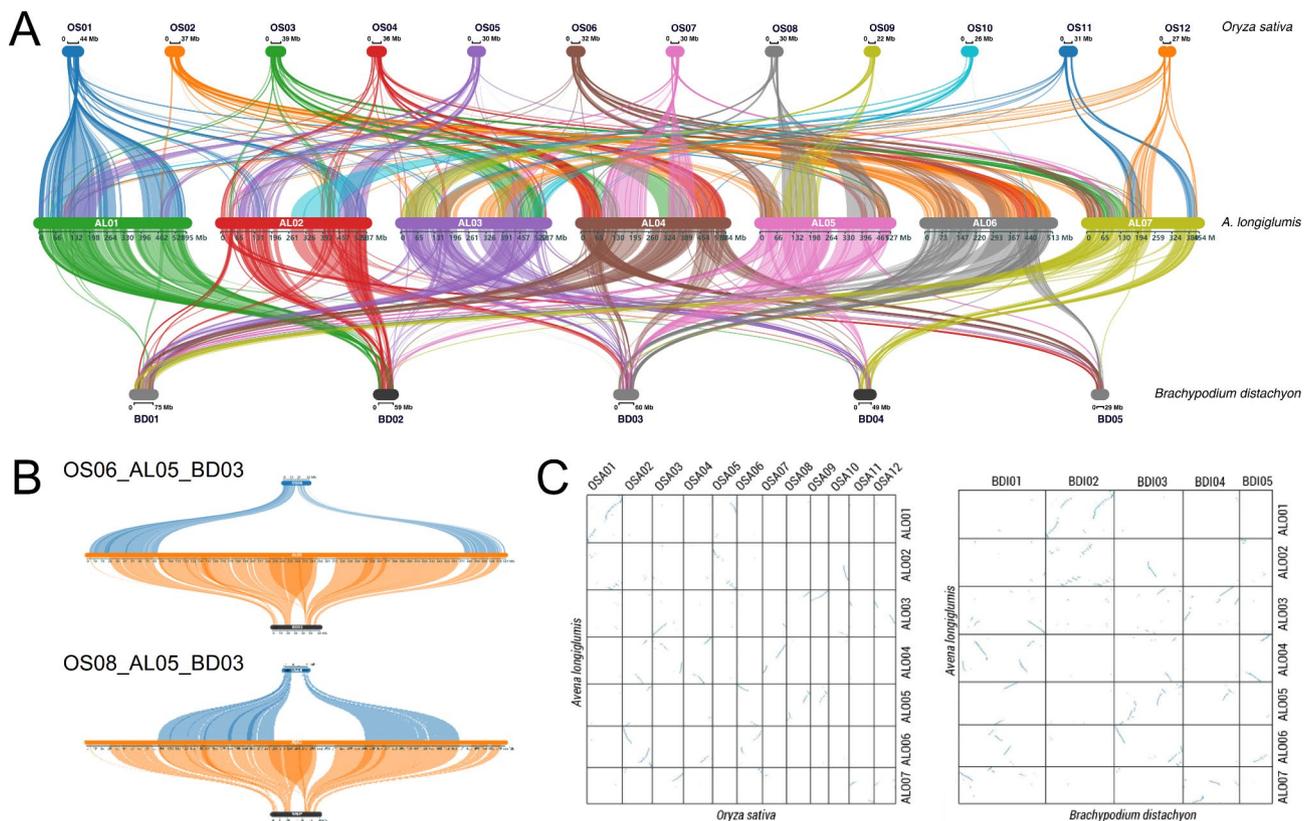


Fig. 1 Deep syntenic relationship of *Oryza sativa* (OSA), *Avena longiglumis* (ALO), and *Brachypodium distachyon* (BDI) chromosomes, drawn to scale, showing detailed conservation of syntenic blocks and the genomic expansion between OSA ($x=12$; 373 Mb), BDI ($x=5$; 271 Mb) and ALO ($x=7$; 3,850 Mb). **A** Syntenic analysis of OSA (top), ALO (middle), and BDI (bottom). Subterminal regions are frequently involved in interspecific evolutionary translocations. **B** Syntenic analyses of chromosomes OS06-AL05-BD03 (top) and OS08-AL05-BD03 (bottom). **C** Dotplots (not to scale) of OSA-ALO (left) and BDI-ALO (right) genomes

BLASTP. With homologous gene data as input, MCS-canX [46] was used to infer homologous blocks involving collinear genes within and between genomes. The maximum gap length between collinear genes along a chromosome region was set to 50 genes [47]. Homology dotplots were constructed using SynVisio [39] to reveal genomic correspondence in ALO, between three *Avena* species, between ancestral grass karyotype (AGK) and seven grass species, and between ALO, OSA, and BDI (Fig. 2, Additional file 2: Table S6).

Fluorescence *in situ* hybridization

Seeds of the hexaploid oat *A. sativa* ($2n=6x=42$) were used for chromosome preparations. The plant materials and probes [AF226603_45bp (labeled with tetra-chloro-fluorescein TET), pAs120a (labeled with biotin) and Ab-T148 (labeled with digoxigenin)] used in this experiment were as reported in Liu et al. [48]. Root tips were fixed in 96% ethanol: glacial acetic acid (3:1) for at least 1.5 h and stored in the fixative at -20°C overnight. An enzyme solution containing 0.2% Cellulase Onozuka R10 (Yakult Pharmaceutical, Tokyo), 2% Cellulase C1184 (Sigma-Aldrich, St Louis, USA), and 3% Pectinase (P4716; Sigma-Aldrich, St Louis, USA) was used to digest root tips for 90 min at 37°C . Finally, root tips were macerated in a drop of 60% acetic acid, and squashed gently under a coverslip.

Fluorescence *in situ* hybridization (FISH) was performed as described by Liu et al. [48] and Schwarzacher

and Heslop-Harrison [49]. The hybridization mixture (stringency 76%), containing 50% formamide, $2 \times$ SSC (Saline Sodium Citrate buffer; 0.3 M NaCl, 0.03 M sodium citrate), 10% dextran sulphate, 0.125% SDS (sodium dodecyl sulphate), 0.125mM EDTA (ethylenediamine-tetraacetic acid), 1 μg sheared salmon sperm DNA, and 100 ng of labeled probes, were applied to each slide. After hybridization at 37°C overnight in a ThermoHybaid HyPro-20, slides were washed in $0.1 \times$ SSC at 42°C . FISH probe hybridization sites were detected via fluorescein isothiocyanate (FITC) conjugated anti-digoxigenin (200 $\mu\text{g}/\text{ml}$; Roche Diagnostics), and streptavidin Alexa Fluor 647 (Molecular Probes, Invitrogen). Slides were counterstained with 4',6-diamidino-2-phenylindole (DAPI; 3 mg/ml)-antifade solution (AF1, Citifluor, London, UK; 50%). FISH images were captured by a Nikon Eclipse 80i epifluorescent microscope fitted with appropriate sets of band-pass filters, a DS-QiMc monochromatic camera, and NIS-Elements v.2.34 (Nikon, Tokyo, Japan). For each metaphase, four 1280×1024 pixel size, single channel (pseudo-colored, yellow, red, green, and blue respectively) images were analyzed using Image J v.1.51j8 (Wayne Rasband, NIH, USA) and superimposed in Photoshop CS6 v.13.0 (Adobe System, San Jose, CA, USA).

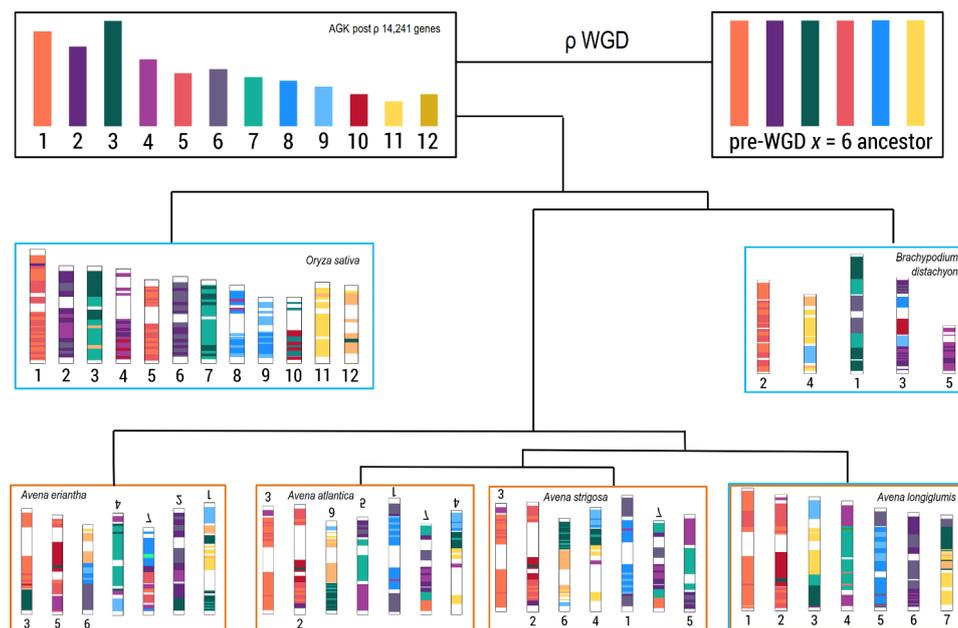


Fig. 2 Reconstruction of ancestral chromosomes for the six species showing conservation of major syntenic blocks from the ancestral grass karyotype (AGK), with fusions and insertions leading to the reduced chromosome numbers (some are upside down to display features of evolutionary conservation). Genes from the ancestral linkage groups are indicated by different colors, with pairs of similar colors representing the pre- ρ whole-genome duplication (WGD)

Results

To understand the chromosomal structural variation, including translocations, duplications, and deletions, we examined the extent of chromosomal rearrangements in Pooideae species by examining gene synteny, gene locations on chromosomes, and the interspersions of non-coding repetitive DNA sequences. From the BOP clade, two species with small genome sizes, *O. sativa* Nipponbare [18] and *B. distachyon* [19], were compared with a new assembly of the diploid oat *A. longiglumis* (Additional file 2, Tables S1 and S2) [36], and three other diploid *Avena* species, *A. atlantica* [20], *A. eriantha* [20], and *A. strigosa* [40]. From a total of 19,954 gene families identified in ALO, about 10% (1,880) were analyzed as orthologous single-copy genes in AAT, AER, AST, as well as BDI and OSA (Additional file 2: Tables S3 and S4).

Conserved synteny and genomic expansion in the BOP clade: *Avena*, rice, and brachypodium

The orthologous genes were used to generate a synteny (McScanX visualized by SynVisio [39, 46]) plot with chromosomes of the diploid species OSA, BDI, and ALO drawn to scale, showing lines linking each conserved group of genes between the species (Fig. 1, Additional file 1: Figure S1). Notably, the dense concentration of synteny lines in the 23–75 Mb chromosomes of OSA and BDI were spread throughout the orthologous 454–595 Mb chromosomes of ALO, apart from the centromeric regions (Fig. 1A). There were few larger gaps along the chromosome arms; and syntenic orthologous genes extended to the subterminal regions of all chromosomes. Figure 1B highlights chromosome AL05, which is largely syntenic to BD03. The distal regions of both chromosome arms are syntenic to OS06 while the central region is syntenic to OS08 (all syntenic relationships between ALO chromosomes shown in Additional file 1: Figure S2). Dotplots (not to scale, Fig. 1C) display large stretches of homologies in straight lines to the chromosome ends (into corners of the plots), supporting that syntenic regions of orthologous genes continue to the ends of chromosomes and emphasizing that there are minimal syntenic regions in broad centromeric regions.

The centromeric regions of ALO have a very low density of single-copy genes, but the *Cereba*-like retrotransposon is abundant (Additional file 1: Figures S3, S4 and Additional file 2: Table S5B) [36]. The regions adjacent to centromeres with lower gene density (large gaps in the SynVisio plots, Fig. 1A) share gene synteny with BDI in some chromosomes, but these genes appear to lack groups of conserved OSA orthologues on chromosomes AL01, AL04, AL05 and AL07 (visualized as no lines in the expanded region around the centromere of ALO chromosomes linked to OSA chromosomes) (Fig. 2, Additional file 1: Figure S2). In the comparison of AL04 and BD01

(Additional file 1: Figure S2N and P), the centromeric gap is translocated. In effect, genes flanking one side of the gap have moved to the other side of centromere between the two species, while the orthologous genes flanking the centromere are missing in OSA03 and OS07.

Ancestral chromosome rearrangement and evolution

The proposed post- ρ AGK [5, 14] has 12 proto-chromosomes, here designated AG01 to AG12, with extensive similarities with rice. We mapped the AGK genes to chromosomes of ALO and five BOP grass species (AAT, AST, AER, BDI, and OSA). The corresponding regions of chromosomes, mosaic synteny blocks, were designed by different colors (Fig. 2, Additional file 2: Table S6). Features of the ancient ρ duplication are shown by chromosome pairs with similar shades of color (Fig. 2), and also shown by dot blots and some syntenic lines being duplicated (Additional file 1: Figure S5). The ancestral synteny blocks defined by shared gene sequences (Fig. 1) were conserved among the analyzed grasses, with distinct rearrangements involving translocations and fusions of syntenic blocks between species. Some rearrangement events are shared between all $x=5$ and $x=7$ species (e.g., the fusion of AG09 and AG11; or AG02 and AG03; both are seen in BDI and *Avena* species) or between the $x=7$ species (AG12 and AG06 giving AL07; Fig. 2, Additional file 2: Table S6). Some evolutionary events associated with the 12 ancestral AGK chromosomes involve fusion and rearrangement of syntenic blocks, but it is notable that three events are characterized by insertion (nesting) of one chromosome into another chromosome. Especially, AL04 has AG07 inserted into AG04, AL06 has much of AG06 inserted into AG02, and AL05 has AG08 inserted into AL06 (Fig. 2).

Chromosomal rearrangements within *Avena*

To evaluate the intraspecific chromosome structure across the genus, we analyzed intragenomic synteny for AAT, ALO, AST, and AER. We found more than 21,000 pairs of collinear genes among ALO-AST, ALO-AER, and AER-AST species pairs (Additional file 2: Table S6). The greater number of rearrangements in AGK with respect to the AGK (Fig. 2) suggests it has the most derived karyotype in *Avena*, while A-genome species (ALO, AST, and AAT) are more primitive. Visualization of syntenic regions between ALO, AST and AER, shows large blocks of conservation between ALO and AST, with much more rearrangements with phylogenetically more distant AER (Fig. 3A). Between ALO and AER, AL01 was collinear with AE03, and AL06 with AE02. Notably, seven evolutionary inter-chromosomal translocations involved large distal domains between 10.64% and 37.24% of the chromosome length, not including centromeric translocations (Fig. 3, Additional file 2: Tables S7, S8, and S9 [50–56]).

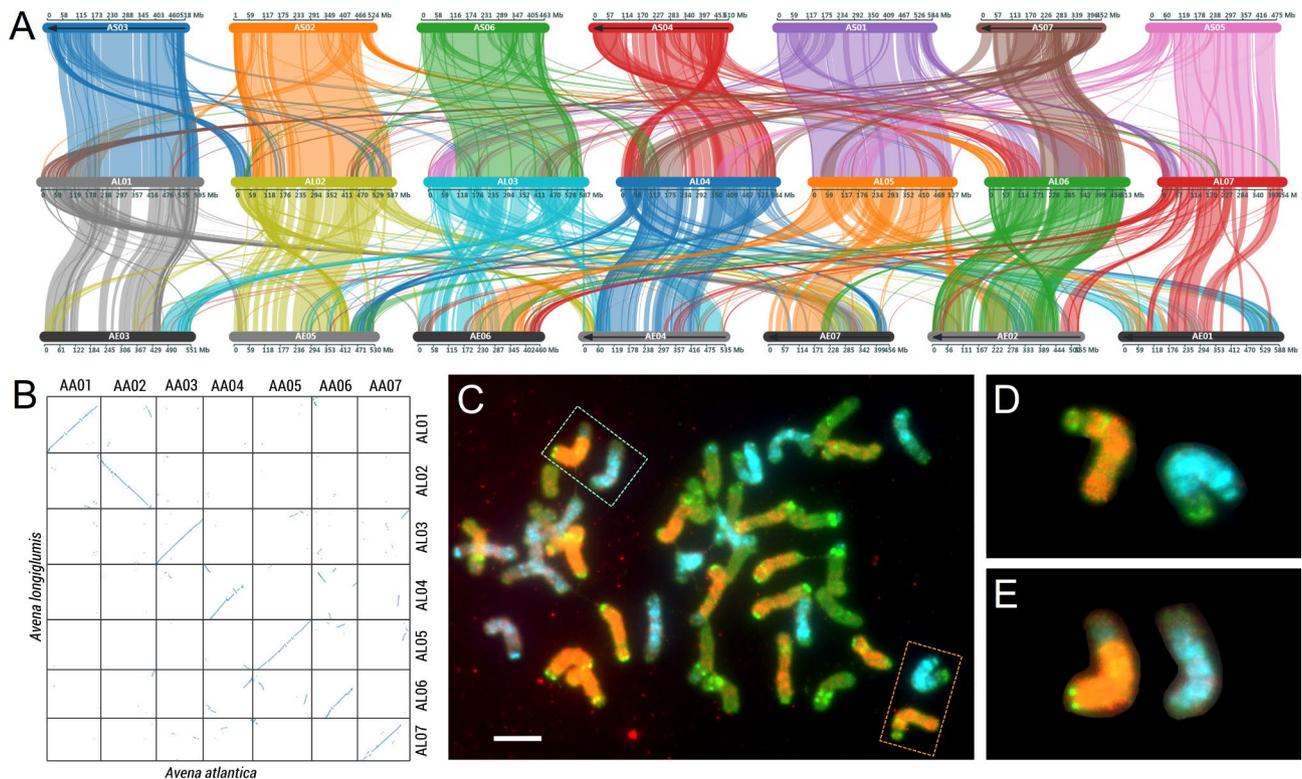


Fig. 3 Syntenic relationship of *Avena strigosa* (AST), *A. longiglumis* (ALO), and *A. eriantha* (AER) genomes

A Syntenic analysis of AST (top), ALO (middle), and AER (bottom). Subterminal regions are frequently involved in inter- and intra-specific evolutionary translocations. **B** Dotplot of AAT-ALO. **C–E** Fluorescent in situ hybridization (FISH) showing evolutionarily recent inter-genomic translocations between *Avena* genomes in the polyploid *A. sativa*. These are similar in extent to those identified by synteny analysis between *Avena* species. **C** FISH karyotype of *A. sativa*. Probes are AF226603_45bp (TET, pseudo-blue) for C genome, pAs120a (biotin, pseudo-red) for A genome, and Ab-T148 (digoxigenin, pseudo-green) for A/D genome. Probes were amplified from *A. longiglumis*. **D** Translocation shown by orange dotted square in Fig. 3C. **E** Translocation shown by green dotted square in Fig. 3C

Figure 3B shows a dotplot comparison of AAT and ALO. Similar distal translocations to those between ALO, AER, and AST are evident, including an intrachromosomal translocation between the ends of chromosomes AA04 and AL05 (Fig. 3B, Additional file 1: Figure S5D).

These distal intragenomic evolutionary chromosomal rearrangements, now identified between diploid *Avena* species, are consistent with the distal nature and size of translocations identified using genome-specific repeat probes in polyploid *Avena* (Fig. 3C–E, see also Liu et al. [48]). In situ hybridization shows the relatively uniform dispersal of many genome-specific repetitive DNA sequences isolated from diploid *Avena* species, consistent with the uniform genomic expansion from the ancestral AGK species with a smaller genome. The hybridization pattern does not show bands of repeats interspersed with repeat-depleted genic regions (Fig. 3C and E). The translocations in hexaploid oat, involving different genomes, have occurred since hybridization and polyploidization but are not accompanied by homogenization of the repeats across the chromosomes, so translocations are revealed by the genome-specific probes. Distal translocations of similar nature and extent were found to occur

during the evolution of the diploid species (Fig. 3), now accompanied by homogenization of the repeats so that they are only detected by the synteny analysis.

Discussion

The presence of thousands of orthologous genes in multiple species has enabled remarkable comparisons of genomes and their evolution over long evolutionary distances [57–59]. Such studies have revealed the conservation of syntenic blocks and their rearrangements as mosaics. Among grasses, polyploidy has played a major part in the earliest (c. 110 Mya, with the ρ event) [4] and most recent (<1 Mya, for example with tetraploid and hexaploid oat and wheat) events in Oryzoideae and Pooideae [11]. In the BOP clade, genomic expansion (Fig. 1) and chromosome reorganization with a number of well-defined events (Fig. 2) complementing wheat [19], indicate that gene duplication, gain, and loss have played a relatively small role. The 3.85 Gb chromosome-scale genome assemblies of *A. longiglumis* and other diploid *Avena* species have facilitated analyses of genomic expansion involving repetitive DNA amplification and homogenization (as discussed in Liu et al. [48] and

Heslop-Harrison and Schwarzacher [60]). These data can be integrated into evolutionary models [6, 11] for comparison of the smaller genomes of rice and Brachypodium to the *Avena* genomes.

Chromosomal block evolution and genomic expansion

The conservation of large syntenic blocks and orthologous relationships between seven ALO chromosomes, twelve chromosomes of OSA, and five of BDI, highlighted chromosomal block evolution with a well-defined number of fusion, translocation and nesting events, but few duplications or deletions (Figs. 1 and 2). Interestingly, certain chromosomes contained many rearrangements and smaller blocks of conserved synteny, while others remain more intact in the three species. The signature of the ρ -WGD event is evident (Additional file 1: Figure S5), and the lack of further large regions of synteny is indicative of a lack of major duplications in the *Avena* lineage. In contrast, the monocotyledons in the Musaceae [26, 61] exhibit evidence of three WGD events, while the PAC-MAD (sister to the BOP) clade contains a tetraploidization event in the maize lineage [62].

Our results revealed that the genomic expansion was uniform along chromosome arms, from telomeres to broader proximal regions around the centromeres, between *A. longiglumis* and rice (10.1-fold smaller) and Brachypodium (15.6-fold smaller) (Fig. 1). The lines of synteny in the dotplots comparing BDI and OSA with ALO (Fig. 1C, Additional file 1: Figure S1) are largely at the same slope (and straight), indicating equal expansion of all syntenic blocks throughout the larger genome. A few individual line segments were curved, indicating greater genomic expansion (spreading out the genes over a longer length) at one end than the other within a syntenic block. This is contrast with *Musa acuminata* and *Ensete glaucum* (two Musaceae species with similar genome sizes; Fig. 8B in Wang et al. [26]), where some lines of synteny were curved and many segments had different slopes. Our results do not support an alternative hypothesis about uneven genomic expansion. For example, we do not see large repeat-domains of integrating hot-spots interspersed between the syntenic blocks, except around the centromeres (supported by the relative uniformity of repeat distribution along chromosome arms in ALO, Additional file 1: Figures S3 and S4). In particular, blocks of tandemly repeated DNAs, often seen as heterochromatin, are not found in *Avena* except around the centromeres. This is in contrast to Triticeae species, which contain large terminal, centromeric, and intercalary blocks of tandem repeats [63]. The expansion of the *Avena* genome with respect to BDI and OSA is largely, although not entirely, accounted for the annotated repetitive element expansion. Specially, 87% of ALO was represented by annotated repeats, compared to 36% in BDI

and 48% in OSA, giving repeat-masked genome sizes of 513, 167 and 203 Mb respectively (Additional file 2: Table S8A and B).

In other Pooideae species, a collinearity analysis between *Lolium perenne* (2,550 Mb) and barley (4,830 Mb) genomes shows even expansion throughout the chromosomes, and no chromosomal segments with markedly stronger sequence expansion or contraction [64]. In contrast, synteny between the three hexaploid wheat (*Triticum aestivum*, $2n=6x=42$) genomes, derived from diploid species *Triticum urartu* (A genome), an *Aegilops speltoides*-related species (B), and *Aegilops tauschii* (D), have uncovered that the smaller D genome, compared to the B genome, shows notable lines of synteny connecting homoeologous genes with gaps or discontinuities (centre of the Circos plot Fig. 1 in El Baidouri et al. [19]), which are not seen between the *Avena* species (Fig. 3A), nor between ALO and BDI or OSA (Fig. 1A, Additional file 1: Figures S2). It is likely that the gaps in the B genome are composed of gene-poor heterochromatic repetitive elements, well-known in the *Aegilops* ancestral species but not seen outside the centromeric region in *Avena* [48].

Massive oligonucleotide pools, synthesizing tens of thousands of synthetic labeled probes, are proving valuable for chromosome evolution studies [65, 66]. Given the extensive gene homology between the BOP grasses shown here, these may enable the design of synthetic chromosome oligonucleotide pools for in situ hybridization to identify syntenic chromosomal blocks and their rearrangements. The use of multiple baits allows isolation of orthologous (and sometimes paralogous) genes from multiple species in the Angiosperm353 projects [67–69] for phylogenetic studies, and it would be exciting if a related probe pool technology could be used to track chromosomal reorganization.

Chromosome evolution during oat diploid speciation

Despite their relatively close relationship, separating between 2 and 10 Mya, we identified substantial rearrangements of syntenic blocks of genes between the four diploid *Avena* species studied here (Figs. 2 and 3). The more diverged AER (designated as C-genome) exhibited more rearrangements than the A-genome species AAT, ALO, and AST. Notably, multiple translocations involving distal regions of chromosome arms were clear from AST to ALO (Fig. 3A, Additional file 1: Figure S1B); and from ALO to AAT (Fig. 3B). The three *Avena* A genomes, with frequent terminal segment translocations, contrast with the Triticeae [19] that show near end-to-end synteny, with no distal arm translocations between $2x$ wheat ancestors (except chromosome 5 A; compare centre of Fig. 1 in El Baidouri et al. [19] with Fig. 3A here). The wheat diploid ancestor phylogenetically separated over

an approximately similar period to the *Avena* species. Thus, we postulate that chromosomal rearrangements have been more active or perhaps more stably in *Avena* compared to wheats (*Triticum*), although both tribes have a conserved chromosome number of $x=7$.

During and following speciation, many of the repetitive elements identified from *Avena* diploid species have become species-specific [48] and, as in Brassicaceae [12], are phylogenetically informative. The sequences have been replaced, lost, amplified, and homogenized along all chromosome arms [48], with little change in *Avena* genome size (cf., genomic expansion in Fig. 1; and *Avena* species in Fig. 3). While broad pericentromeric regions are reservoirs for accumulation of a medley of TEs [26, 60, 70].

In situ hybridization using genome (species)-specific repeat probes shows that in the hexaploid *A. sativa* (Fig. 3C–E), many chromosomes contain intergenomic translocations between chromosomes of diploid genomes [71, 72], involving the terminal 10.64–37.24% of chromosome arms (Additional file 2: Table S7). Our analysis of conserved gene synteny (Fig. 3A and B) revealed that the four diploid *Avena* species (ALO, AST, AAT, and AER) contain multiple terminal translocations between chromosomes. Notably, the terminal rearrangements involved more than just repetitive DNAs, as is the case in maize (The P53 knob [73] or rye (pSc250 tandem repeat) [63], and include many genes in the synteny [74, 75]. The hexaploid result shows that distal translocation events in *Avena* continue to occur post-polyploidization, between chromosomes of different species origin (Fig. 3C–E).

Both genomic expansion and chromosomal rearrangement have occurred during evolution of *Avena* from a proposed AGK similar to rice, without further rounds of polyploidy. Chromosomal structural variation is extensive, and may restrict hybridization and lead to reproductive isolation. While a key feature of speciation, this phenomenon restricts crossing in breeding programs to exploit wider germplasm pools. Chromosome structural variation is increasingly recognized as a factor controlling complex traits in livestock [76] and crop plants [77], and must be discerned as a part of the pangenome [26, 78]. The 10-fold to 15-fold genomic expansion involving relatively uniform interspersions of genes with repetitive DNAs throughout chromosome arms, along with changes in the size of gene-depleted broad centromeric regions, may also contribute to modulation of gene expression, and perhaps reproductive isolation, although meiotic pairing can compensate for substantial genome-size differences [79].

Insight into the extent and nature of chromosomal rearrangements and genomic expansion in the pangenome is critical for identifying the processes of evolution and speciation. Beyond the level of gene sequences,

this information can inform studies on biodiversity, and contribute to the exploitation of diversity present in the common gene pool across grasses through precision breeding. Pangenomic resources will allow us to increase power in genome editing and synthetic biology [80], reduce costs by saving resources required for extensive phenotypic selections [81], speed up the process of genetic improvement [82], and realize the genetic gains per unit time with high precision [83]. Therefore, the knowledge gained in the genomic expansion and reorganization of the BOP clade can be rapidly transferred to exploit biodiversity and widen gene pools available to the genomic-assisted breeding programs for future crops [84, 85].

Abbreviations

AAT	<i>Avena atlantica</i>
AER	<i>A. eriantha</i>
AGK	ancestral grass karyotype
ALO	<i>A. longiglumis</i>
AST	<i>A. strigosa</i>
BDI	<i>Brachypodium distachyon</i>
BOP	Bamusoideae, Oryzoideae, and Pooideae
DAPI	4',6-diamidino-2-phenylindole
EDTA	ethylenediamine-tetraacetic acid
FISH	Fluorescence in situ hybridization
FITC	fluorescein isothiocyanate
Hi-C	high-throughput chromatin conformation capture
Mb	Megabases
Mya	million years ago
OSA	<i>Oryza sativa</i>
NCBI	National Center for Biotechnology Information
PACMAD	Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae, and Danthonioideae
SDS	sodium dodecyl sulphate
SSC	saline sodium citrate
WGD	whole-genome duplication

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-023-04644-7>.

Supplementary Material 1

Acknowledgements

We thank National Plant Germplasm System (NGPS) of USDA-ARS for providing seeds, Kai Ouyang for expertise in the orthogroup analysis, Grandomics Biosciences Co., Ltd. (Wuhan, China) for genome sequencing support, and Huawei Elastic Cloud Server (Guiyang, China) for on-demand computational resources. No conflict of interest declared.

Authors' contributions

QL and JSHH designed the research. MZL, ZWW, GX, and YSY conducted experiments. LHY, MZL, ZWW, and GX assembled and annotated the genome. LHY and GX performed bioinformatics analyses. QL, TS, TYT and JSHH wrote the manuscript. All authors read and approved the final manuscript.

Funding

This research was funded by National Natural Science Foundation of China (32070359, 32370402), Guangdong Basic and Applied Basic Research Foundation (2021A1515012410), Guangdong Provincial Special Fund for Natural Resource Affairs on Ecology and Forestry Construction (GDZZDC20228704), Chinese Academy of Sciences (CAS) President's International Fellowship Initiative (2024PVA0028), Global Challenges

Research Foundation for Global Agricultural and Food Systems Research (BB/P02307X/1), and Sciences Innovative Training Programs for Undergraduates of Chinese Academy of Sciences (KCJH-80107-2023-148).

Data Availability

The raw sequencing reads of Nanopore, Hi-C and Illumina sequencing data that were used for the genome assembly have been deposited in the NCBI Sequence Read Archive with accession number SRR19279519-SRR19279520 (Nanopore), SRR19279522-SRR19279531 (Nanopore), SRR19279511-SRR19279517 (Hi-C), SRR19279521, SRR19279532-SRR19279533 (Hi-C), and SRR19279518 (Survey data) under BioProject accession number PRJNA838431. The assembly data have been deposited in the NCBI under the BioProject ID PRJNA956334. The chromosomal assembly and annotation are also available on Figshare with the identifier <https://doi.org/10.6084/m9.figshare.19130429.v2>.

Declarations

Ethics approval and consent to participate

We sampled the diploid oat *Avena longiglumis* in this study from publicly available seed sources. The plant samples and experimental research comply with relevant institutional, national, and international guidelines and legislation.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 31 August 2023 / Accepted: 29 November 2023

Published online: 08 December 2023

References

1. APG (Angiosperm Phylogeny Group). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants. *Bot J Linn Soc.* 2016;181(1):1–20.
2. Heslop-Harrison JS, Schwarzacher T, Liu Q. Polyploidy: its consequences and enabling role in plant diversification and evolution. *Ann Bot.* 2023;131(1):1–10.
3. Stull GW, Pham KK, Soltis PS, Soltis DE. Deep reticulation: the long legacy of hybridization in vascular plant evolution. *Plant J.* 2023;114(4):743–66.
4. McKain MR, Tang HB, McNeal JR, Ayyampalayam S, Davis JL, dePamphilis CW, et al. A phylogenomic assessment of ancient polyploidy and genome evolution across the Poales. *Genome Biol Evol.* 2016;8(4):1150–64.
5. Murat F, Xu JH, Tannier E, Abrouk M, Guilhot N, Pont C, et al. Ancestral grass karyotype reconstruction unravels new mechanisms of genome shuffling as a source of plant evolution. *Genome Res.* 2010;20(11):1545–57.
6. Wang XY, Wang JP, Jin DC, Guo H, Lee TH, Liu T, et al. Genome alignment spanning major Poaceae lineages reveals heterogeneous evolutionary rates and alters inferred dates for key evolutionary events. *Mol Plant.* 2015;8(6):885–98.
7. Alix K, Gérard PR, Schwarzacher T, Heslop-Harrison JS. Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. *Ann Bot.* 2017;120(2):183–94.
8. Liu Q, Peterson PM, Ge XJ. Phylogenetic signals in the realized climate niches of Chinese grasses (Poaceae). *Plant Ecol.* 2011;212(1):1733–46.
9. Vanneste K, Baele G, Maere S, Van de Peer Y. Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-Paleogene boundary. *Genome Res.* 2014;24(8):1334–7.
10. Soreng RJ, Peterson PM, Romaschenko K, Davidse G, Teisher JK, Clark LG, et al. A worldwide phylogenetic classification of the Poaceae (Gramineae) II: an update and a comparison of two 2015 classifications. *J Syst Evol.* 2017;55(4):259–90.
11. Bellec A, Dia Sow M, Pont C, Mardoc E, Duchemin W, Armisen D, et al. Tracing 100 million years of grass genome evolutionary plasticity. *Plant J.* 2023;114(6):1243–66.
12. Lysak MA, Mandáková T, Schranz ME. Comparative paleogenomics of crucifers: ancestral genomic blocks revisited. *Curr Opin Plant Biol.* 2016;30(1):108–15.
13. Cao Y, Zhao KL, Xu JX, Wu L, Hao FY, Sun MP, et al. Genome balance and dosage effect drive allopolyploid formation in *Brassica*. *Proc Natl Acad Sci USA.* 2023;120(14):e2217672120.
14. Murat F, Armero A, Pont C, Klopp C, Salse J. Reconstructing the genome of the most recent common ancestor of flowering plants. *Nat Genet.* 2017;49(4):490–6.
15. Fu YB. Characterizing chloroplast genomes and inferring maternal divergence of the *Triticum-Aegilops* complex. *Sci Rep.* 2021;11(1):15363.
16. Gordon SP, Contreras-Moreira B, Levy JJ, Djamei A, Czedik-Eysenberg A, Tartaglio VS, et al. Gradual polyploid genome evolution revealed by pan-genomic analysis of *Brachypodium hybridum* and its diploid progenitors. *Nat Commun.* 2020;11(1):3670.
17. International Brachypodium Initiative. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature.* 2010;463(7282):763–8.
18. Ouyang S, Zhu W, Hamilton J, Lin HN, Campbell M, Childs K, et al. The TIGR rice genome annotation resource: improvements and new features. *Nucleic Acids Res.* 2007;35(D1):D883–7.
19. El Baidouri M, Murat F, Veyssié M, Molinier M, Flores R, Burlot L, et al. Reconciling the evolutionary origin of bread wheat (*Triticum aestivum*). *New Phytol.* 2017;213(3):1477–86.
20. Maughan PJ, Lee R, Walstead R, Vickerstaff RJ, Fogarty MC, Brouwer CR, et al. Genomic insights from the first chromosome-scale assemblies of oat (*Avena* spp.) diploid species. *BMC Biol.* 2019;17(1):92.
21. Tørresen OK, Star B, Mier P, Andrade-Navarro MA, Bateman A, Jarnot P, et al. Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. *Nucl Acids Res.* 2019;47(21):10994–1006.
22. Kress WJ, Soltis DE, Kersey PJ, Wegrzyn JL, Leebens-Mack JH, Gostel MR, et al. Green plant genomes: what we know in an era of rapidly expanding opportunities. *Proc Natl Acad Sci USA.* 2022;119(4):e2115640118.
23. Belser C, Istace B, Denis E, Dubarry M, Baurens FC, Falentin C, et al. Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat Plants.* 2018;4(11):879–87.
24. Amarasinghe SL, Su S, Dong XY, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol Evol.* 2020;21(1):30.
25. Perumal S, Koh CS, Jin L, Buchwaldt M, Higgins EE, Zheng CF, et al. A high-contiguity *Brassica nigra* genome localizes active centromeres and defines the ancestral *Brassica* genome. *Nat Plants.* 2020;6(8):929–41.
26. Wang ZW, Rouard M, Biswas MK, Droc G, Cui DL, Roux N, et al. A chromosome-level reference genome of *Ensete glaucum* gives insight into diversity, chromosomal and repetitive sequence evolution in the Musaceae. *GigaScience.* 2022;11(1):giac027.
27. Ou SJ, Jiang N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* 2018;176(2):1410–22.
28. Ellinghaus D, Kurtz S, Willhoeft U. LTRharvest, an efficient and flexible software for *de novo* detection of LTR retrotransposons. *BMC Bioinformatics.* 2008;9(1):18.
29. Ou SJ, Jiang N. LTR_FINDER_parallel: parallelization of LTR_FINDER enabling rapid identification of long terminal repeat retrotransposons. *Mob DNA.* 2019;10(1):48.
30. Shi M, Liang C. Generic repeat finder: a high-sensitivity tool for genome-wide *de novo* repeat detection. *Plant Physiol.* 2019;180(4):1803–15.
31. Su W, Peterson T. TIR-learner, a new ensemble method for TIR transposable element annotation, provides evidence for abundant new transposable elements in the maize genome. *Mol Plant.* 2016;12(3):447–60.
32. Xiong WW, He LM, Lai JS, Dooner HK, Du CJ. HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *Proc Natl Acad Sci USA.* 2014;111(28):10263–8.
33. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA.* 2020;117(17):9451–7.
34. Smit AF, Hubley R, Green P. RepeatModeler Open-1.0. 2008–2015. Institute for Systems Biology: Seattle, USA, 2015.
35. Zhang RG, Wang ZX, Ou SJ, Li GY. TESorter: lineage-level classification of transposable elements using conserved protein domains. *bioRxiv [Preprint]* 2019; <https://www.biorxiv.org/content/10.1101/800177v1> (accessed 17 August 2023).

36. Liu Q, Yuan HY, Li MZ, Wang ZW, Cui DL, Ye YS, et al. Chromosome-scale genome assembly of the diploid oat *Avena longiglumis* reveals the landscape of repetitive sequences, genes and chromosome evolution in grasses. *bioRxiv* [Preprint]. 2022. <https://doi.org/10.1101/2022.1102.1109.479819>. (accessed 17 August 2023).
37. Tomás D, Rodrigues J, Varela A, Veloso MM, Viegas W, Silva M. Use of repetitive sequences for molecular and cytogenetic characterization of *Avena* species from Portugal. *Int J Mol Sci*. 2016;17(2):203.
38. International Wheat Genome Sequencing Consortium (IWGSC). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*. 2014;345(6194):1251788.
39. Bandi V, Gutwin C. "Interactive exploration of genomic conservation" in Proceedings of the 46th Graphics Interface Conference on Proceedings of Graphics Interface 2020. Canadian Human-Computer Communications Society, Waterloo, Canada, 2020; pp. 1–10.
40. Li Y, Leveau A, Zhao Q, Feng Q, Lu HY, Miao JS, et al. Subtelomeric assembly of a multi-gene pathway for antimicrobial defense compounds in cereals. *Nat Commun*. 2021;12(1):2563.
41. Goodstein DM, Shu SQ, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 2012;40(D1):D1178–86.
42. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;20(1):238.
43. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12(1):59–60.
44. Tang HB, Lyons E, Pedersen B, Schnable JC, Paterson AH, Freeling M. Screening syntenic blocks in pairwise genome comparisons through integer programming. *BMC Bioinformatics*. 2011;12(1):102.
45. Glover N, Dessimoz C, Ebersberger I, Forslund SK, Gabaldón T, Huerta-Cepas J, et al. Advances and applications in the quest for orthologs. *Mol Biol Evol*. 2019;36(10):2157–64.
46. Wang YP, Tang HB, DeBarry JD, Tan X, Li JP, Wang XY, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012;40(7):e49.
47. Wang J, Yu JX, Sun PC, Li YC, Xia YY, Liu YZ, et al. Comparative genomics analysis of rice and pineapple contributes to understand the chromosome number reduction and genomic changes in grasses. *Front Genet*. 2016;7(1):174.
48. Liu Q, Li XY, Zhou XY, Li MZ, Zhang FJ, Schwarzacher T, et al. The repetitive DNA landscape in *Avena* (Poaceae): chromosome and genome evolution defined by major repeat classes in whole-genome sequence reads. *BMC Plant Biol*. 2019;19(1):226.
49. Schwarzacher T, Heslop-Harrison P. Practical. *in situ* Hybridization. BIOS Scientific Publishers: Oxford, UK, 2000.
50. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, et al. The COG sdatabase: an updated version includes eukaryotes. *BMC Bioinformatics*. 2003;4(1):41.
51. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The pfam protein families database in 2019. *Nucleic Acids Res*. 2019;47(D1):D427–32.
52. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res*. 2019;47(D1):D309–14.
53. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R et al. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*. 2004;32(D1):D258–61 (2004).
54. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, Diseases and Drugs. *Nucleic Acids Res*. 2017;45(D1):D353–61.
55. Tian F, Yang DC, Meng YQ, Jin JP, Gao G. PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Res*. 2020;48(D1):D1104–13.
56. Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol Biofuels*. 2013;6(1):41.
57. Salse J. Deciphering the evolutionary interplay between subgenomes following polyploidy: a paleogenomics approach in grasses. *Am J Bot*. 2016;103(7):1167–74.
58. Zhao T, Zwaenepoel A, Xue JY, Kao SM, Li Z, Schranz ME, et al. Whole-genome microsynteny-based phylogeny of angiosperms. *Nat Commun*. 2021;12(1):3498.
59. Mutwil M, Fernie AR. Ancestral genome reconstruction for studies of the green lineage. *Mol Plant*. 2023;16(4):657–9.
60. Heslop-Harrison JS, Schwarzacher T. Organisation of the plant genome in chromosomes. *Plant J*. 2011;66(1):18–33.
61. D'Hont A, Denoeud F, Aury JM, Baurens FC, Carreel F, Garsmeur O, et al. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature*. 2012;488(7410):213–7.
62. Estep MC, DeBarry JD, Bennetzen JL. The dynamics of LTR retrotransposon accumulation across 25 million years of panicoid grass evolution. *Heredity* (Edinb). 2013;110(2):194–204.
63. Vershinin AV, Alkhimova AG, Heslop-Harrison JS, Potapova TA, Omelianchuk N. Different patterns in molecular evolution of the *Triticum*. *Hereditas*. 2001;135(2–3):153–60.
64. Nagy I, Veekman E, Liu C, Van Bel M, Vandepoel K, Jensen CS, et al. Chromosome-scale assembly and annotation of the perennial ryegrass genome. *BMC Genomics*. 2022;23(1):505.
65. Agrawal N, Gupta M, Banga SS, Heslop-Harrison JS. Identification of chromosomes and chromosome rearrangements in crop brassicas and *Raphanus sativus*: a cytogenetic toolkit using synthesized massive oligonucleotide libraries. *Front Plant Sci*. 2020;11(1):598039.
66. Shi PY, Sun HJ, Liu GQ, Zhang X, Zhou JW, Song RR, et al. Chromosome painting reveals inter-chromosomal rearrangements and evolution of subgenome D of wheat. *Plant J*. 2022;112(1):55–67.
67. Johnson MG, Pokorny L, Dodsworth S, Botigué LR, Cowan RS, Devault A, et al. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Syst Biol*. 2019;68(4):594–606.
68. Baker WJ, Dodsworth S, Forest F, Graham SW, Johnson MG, McDonnell A, et al. Exploring Angiosperms353: an open, community toolkit for collaborative phylogenomic research on flowering plants. *Am J Bot*. 2021;108(7):1059–65.
69. Yu F, Zhao XW, Chai J, Ding XE, Li XT, Huang YJ, et al. Chromosome-specific painting unveils chromosomal fusions and distinct allopolyploid species in the *Saccharum* complex. *New Phytol*. 2022;233(4):1953–65.
70. Nishihara H. Transposable elements as genetic accelerators of evolution: contribution to genome size, gene regulatory network rewiring and morphological innovation. *Genes Genet Syst*. 2019;94(6):269–81.
71. Katsiotis A, Loukas M, Heslop-Harrison JS. Repetitive DNA, genome and species relationships in *Avena* and *Arrhenatherum*. *Ann Bot*. 2000;86(6):1135–42.
72. Jiang WX, Jiang CZ, Yuan WG, Zhang MJ, Fang ZJ, Li Y, et al. A universal karyotypic system for hexaploid and diploid *Avena* species brings oat cytogenetics into the genomics era. *BMC Plant Biol*. 2021;21(1):213.
73. Birchler JA, Han FP. McClintock's unsolved chromosomal mysteries: parallels to common rearrangements and karyotype evolution. *Plant Cell*. 2018;30(4):771–9.
74. Yang XF, Gao SH, Guo L, Wang B, Jia YY, Zhou J, et al. Three chromosome-scale *Papaver* genomes reveal punctuated patchwork evolution of the morphinan and noscapine biosynthesis pathway. *Nat Commun*. 2021;12(1):6030.
75. Ramin M, Fant P, Huhtanen P. The effects of gradual replacement of barley with oats on enteric methane emissions, rumen fermentation, milk production, and energy utilization in dairy cows. *J Dairy Sci*. 2021;104(5):5617–30.
76. Li R, Gong M, Zhang XM, Wang F, Liu ZY, Zhang L, et al. A sheep pangenome reveals the spectrum of structural variations and their effects on tail phenotypes. *Genome Res*. 2023;33(3):1–15.
77. Yan HD, Sun M, Zhang ZR, Jin YR, Zhang AL, Lin C, et al. Pangenomic analysis identifies structural variation associated with heat tolerance in pearl millet. *Nat Genet*. 2023;55(3):507–18.
78. Coletta RD, Qiu YJ, Ou SJ, Hufford MB, Hirsch CN. How the pan-genome is changing crop genomics and improvement. *Genome Biol*. 2021;22(1):3.
79. Majka J, Glombik M, Doležalová A, Kneřová J, Ferreira MTM, Zwierzykowski Z, et al. Both male and female meiosis contribute to non-mendelian inheritance of parental chromosomes in interspecific plant hybrids (*Lolium × Festuca*). *New Phytol*. 2023;238(2):624–36.
80. Lv RL, Gou XW, Li N, Zhang ZB, Wang CY, Wang RS, et al. Chromosome translocation affects multiple phenotypes, causes genome-wide dysregulation of gene expression, and remodels metabolome in hexaploid wheat. *Plant J*. 2023;115(6):1564–82.
81. Chevegatti-Gianotto A, de Abreu HMC, Arruda P, Bespalhok Filho JC, Burnquist WL, Creste S, et al. Sugarcane (*Saccharum × officinarum*): a reference study for the regulation of genetically modified cultivars in Brazil. *Trop Plant Biol*. 2011;4(1):62–89.
82. Racedo J, Gutiérrez L, Perera MF, Ostengo S, Pardo EM, Cuenya MI, et al. Genome-wide association mapping of quantitative traits in a breeding population of sugarcane. *BMC Plant Biol*. 2016;16(1):142.

83. Rönspies M, Dorn A, Schindele P, Puchta H. CRISPR-Cas mediated chromosome engineering for crop improvement and synthetic biology. *Nat Plants*. 2021;7(5):566–73.
84. Rasheed A, Mujeeb-Kazi A, Ogbonnaya FC, He ZH, Rajaram S. Wheat genetic resources in the post-genomics era: promise and challenges. *Ann Bot*. 2018;121(4):603–16.
85. Song B, Ning WD, Wei D, Jiang MY, Zhu K, Wang XW, et al. Plant genome resequencing and population genomics: current status and future prospects. *Mol Plant*. 2023;16(8):1252–68.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.