Initial description of the Genome of *Aeluropus littoralis*, a halophile grass

Seyyed Hamidreza Hashemi
GABIT
Mojhdeh Arab
GABIT
Behnaz Dolatabadi
Leibniz-Institut fur Pflanzengenetik und Kulturpflanzenforschung Gatersleben
Yi-Tzu Kuo
Leibniz-Institut fur Pflanzengenetik und Kulturpflanzenforschung Gatersleben
Mariana Alejandra Baez
Leibniz-Institut fur Pflanzengenetik und Kulturpflanzenforschung Gatersleben
Axel Himmelbach
Leibniz-Institut fur Pflanzengenetik und Kulturpflanzenforschung Gatersleben
Ghorbanali Nematzadeh
GABIT
Seyed Ali Mohammad Mirmohammady Maibody
Isfahan University of Technology
Thomas Schmutzer
Martin-Luther-Universitat Halle-Wittenberg
Michael Mälzer
Leibniz-Institut fur Pflanzengenetik und Kulturpflanzenforschung Gatersleben
Thomas Altmann
Leibniz-Institut fur Pflanzengenetik und Kulturpflanzenforschung Gatersleben
Markus Kuhlmann (kuhlmann@ipk-gatersleben.de)
Leibniz-Institute of plant genetics and crop plant research (IPK) https://orcid.org/0000-0003-3104-0825

Research article

Keywords: Aeluropus littoralis, halophile, genome, genome size, repetitive elements

Posted Date: December 1st, 2020

DOI: https://doi.org/10.21203/rs.3.rs-53457/v2

Abstract

Background: The use of wild plant species or their halophytic relatives has been considered in plant breeding programs to improve salt and drought tolerance in crop plants. *Aeluropus littoralis* serves as halophyte model for identification and isolation of novel stress adaptation genes. This species is described as perennial monocot grass. *A. littoralis* grows in damp or arid areas, often salt-impregnated places and waste land in cultivated areas. *A. littoralis* can survive where the water salinity is periodically high and tolerate high salt concentrations in the soil up to 1100 mM sodium chloride. Therefore, it serves as valuable genetic resource to understand molecular mechanisms of stress-responses in monocots. The knowledge can potentially be used for improving tolerance to abiotic stresses in economically important crops. Several morphological, anatomical, ecological, and physiological traits of *A. littoralis* have been investigated so far and also the transfer of stress related genes to other species resulted in enhanced stress resistance. After watering with salt water the grass is able to excrete salt via its salt glands. Meanwhile, a number of ESTs (expressed sequence tag), genes and promoters induced by the salt and drought stresses were isolated, sequenced and annotated at a molecular level.

Results: Here we describe the genome sequence and structure of *A. littoralis* analyzed by whole genome sequencing and histological analysis. The chromosome number was determined to be 20 (2n = 2X = 20), absence of B chromsomes shown, and the genome size calculated to be 354 Megabasepairs.

Conclusions: This genomic information provided here, will support the functional investigation and application of novel genes improving salt stress resistance in crop plants.

Background

The use of wild plant species or their halophytic relatives has been considered in plant breeding programs to improve salt and drought tolerance in crop plants [1]. *Aeluropus littoralis* [2] is a monocot belonging to the gramineae (poaceae) family, subfamiliy *Chloridoideae* [3], also referred to as “indian walnut” and first described 1764 by Antoine Gouan (**Figure 1 A**). It serves as halophyte model for identification and isolation of novel stress adaptation genes. This species is described as perennial grass with an estimated haploid genome of 349-8,232 Mbp [4-6] and it possess a C_4 mechanism for carbon fixation [7, 8] with Kranz anatomy and a Mediterranean, Irano-Turanian Chlorotype [9] isolated from their natural habitat. An early study [10] described the influence of salt, shifting the C_3 metabolism toward C_4 metabolism. Such change was also reported lately for other halophile plants [11], but is still debated as such mechanism was not described for any other poacea yet. A salt induced change from C_3 to CAM metabolismus is also a frequently observed strategy of plants to cope with high levels of salt [12, 13]. *A. littoralis* is widely distributed and can be found in northern Africa, in temperate and tropical areas of Asia, southern and south eastern part of Europe. *A. littoralis* grows in damp or arid areas, often salt-impregnated places and waste land in cultivated areas [14]. *A. littoralis* is primarily found in desert regions and regions with high soil salinity due to flooding and can survive where the water salinity is periodically high [15] and tolerate high salt concentrations in the soil up to 1100 mM sodium chloride [16]. The plant is able to secrete salt via its salt glands leading to formation of salt crystals on the leaf surface [8, 17]. From the economic point of view the plants are important for reclaiming salinized agricultural and rangeland, they are used for sand fixation and grow on pastures. Particularly in developing countries [18] they are extensively used as fodder crop. The grass is also capable of vegetative reproduction through rhizome growth after monsoon rains and can produce numerous flowers and seeds from April to October [19]. Due to its high salt tolerance *A. littoralis* serves as valuable genetic resource to understand molecular mechanisms of stress-responses in monocots [20]. These knowledge can potentially be used for improving tolerance to abiotic stresses in economically important crops [21]. Several morphological, anatomical, ecological, and physiological traits of *A. littoralis* have been investigated so far [16, 22, 23] and also the transfer of stress related genes to other species resulted in enhanced stress resistance [24-28].

This plant can grow up to a height of 30 cm. The leaves are distichous and leaf sheaths are longer than adjacent culm internode (**Figure 1 B**). The leaf blades are 1-5 cm in length and 1-2 mm wide. They appear stiff and glaucous, while the leaf surface is ribbed. The inflorescence is composed of two to twelve racemes borne along a central axis. The central

inflorescence axis is 1–4 cm long and the solitary spikelets are packed on the broadside of the rachis [29]. They are on a bilateral false spike and can be termed two-rowed. The spikelets comprise of six to nine fertile florets with diminished florets at the apex.

As several contradicing genome features are found in the literature (genome size, chromosome number, presence of B chromosomes), these points were revisited here in addition to the genome sequence information.

Cytogenetically, for the chromosome number of *A. littoralis* a variation between 2n = 2x = 10 and 2n = 2x = 14 was previously reported and deposited in different [4, 6] Chromosome databases [30]. Likewise, the genome size varies from 342 Mb [4, 6] to 8215 Mbp [5]. To validate the chromosome number of the used material we performed chromosome countings. Futhermore, we analysed the size, structure and composition of the *A. littoralis* genome. Using a High throughput sequencing approach on the genome sequence we present a first assembly of the genome sequence. Meanwhile, a number of ESTs (expressed sequence tag), genes and promoters induced by the salt and drought stresses were isolated, sequenced and annotated at a molecular level [1, 14]. Here we describe the genome sequence and structure of *A. littoralis* analyzed by whole genome sequencing and histological analysis. This genomic information will support the functional investigation and application of novel genes improving stress resistance in crop plants.

Results

For the genome analysis *Aeluropus littoralis* leaf tissue grown under greenhouse conditions, was used. Seeds were collected from Isfahan province in central Iran. This region experiences a moderate and dry climate with temperature ranging between 10.6 °C. and 40.6 °C. The annual rainfall in this region on an average has been reported as 116.9 mm and can be considered super arid (desert) climate. Salt stress condition was applied by watering with 1M NaCl solution instead of tap water under greenhouse conditions. When the plants are exposed to high amounts of salt water (e.g. 1M NaCl) they start to develop salt glands and extrude the salt in crystals on the leaves [8, 17]. **Figure 2A** shows the formed salt crystals on the leave surface. Crystals of cubic shape are formed on the adaxial and abaxial side of the leaf at the salt glands. As very early reports indicate a shift from C_3 to C_4 carbon fixation mechanism, the leaf ultra-structure under control and salt condition was analysed with light and transmission electron microscopy. New leaves developed at control and salt watering conditions from three independent plants each, were analyzed. Under both conditions a Kranz anatomy structure (**Figure 2B**) was found: the enlarged bundle sheath (BS) cells surround the veins and the BS cells are then surrounded by mesophyll (M) cells. Interestingly, when stained with methylenblue/azur II [31] the bundle sheath cells appear darker under salt stress conditions. This might indicate an accumulation of acidic components under salt stress. The bundle sheath cells appears more closed and filled with thylacoids. The ultrastructural analysis shows the increase of tylacoid staples and more spacings between the staples (**Figure 2C**), leading to a higher volume. Whether this indicates a higher activity or a disintegration of chloroplasts coupled with repair mechanism remains to be solved.

Size and structure of *Aeluropus littoralis* genome

In order to support our whole genome sequencing data, we were addressing the question of the *A. littoralis* genome structure. Therefore the nuclear genome size was estimated by flow cytometry (**Figure 3A**) using *Raphanus sativus* cv. 'Voran' (Genebank Gatersleben accession number: RA 34; 2C = 1.11 pg) as reference [32]. The relative fluorescence intensities of around 7,000 to 10,000 events (nuclei) per sample were measured and the absolute DNA amounts of samples were calculated based on the values of the G1 peak means. The DNA content of diploid *A. littoralis* was estimated to be 0.724 ± 0.01 pg/2C (354 Mbp/1C) and is therewith only slightly bigger than previously reported [7]. To validate the chromosome number, karyotyping was performed on mitotic chromosome spreads. The chromosome number was determined to be 20 (2n = 2X = 20, **Figure 3B**). However, ocassionally also metaphases with 21 or 22 chromatin units were found. To analyse if these additional chromatin units resulting from satellites being located distally from the corresponding chromosomes or indeed B-chromosomes, as it was sporadically reported (29), fluorescence *in situ* hybridization (FISH) with the *Arabibidopsis*-type telomere repeat was performed. The resulting hybridization pattern indicates that the increased number of chromatin units are

consequences of extended nucleolus organizing regions (NORs) leading to large gaps between the satellites and the corresponding chromosomes (read arrows, Figure 3C).

Genome Sequencing Information

The genome of *Aeluropus littoralis* was sequenced using a whole-genome sequencing approach (WGS) on Illumina's HiSeq 2500 system. In total, 125 million paired end (PE) reads were produced, reaching genome coverage after quality trimming of approximately 62-fold. This data was sufficient to perform a *de novo* assembly to construct the first available genomic reference for the *A. littoralis* species. The constructed genome sequence reached a total size of ~300 Mbp, which is 87.7 % of the estimated genome size of 342 Mbp [6] or 84,7% of 354 Mbp, estimated here. The assembly consists of 182,747 contigs with a N50 contig length of 3.6 kb. The constructed genomic resource was used for gene prediction and was complemented by a functional annotation of genes. In total, 15,916 gene models were predicted for *A. littoralis* and 87.5% of them could be assigned with a function based on sequence similarity to known genes. (**Table 1, Suppl. Figure 1**) The BUSCO (Simão *et al.*, 2015) analysis revealed ~63 % of the proposed BUSCO genes. In an additional analysis we performed a *de novo* gene prediction using Augustus [33] and allowed to predict also partial genes (which are filtered out in our main gene set). Here, 108,264 putative 'genes' were predicted which we refer as unfiltered gene set. Running BUSCO in this unfiltered set resulted in a completeness ratio of 89%. But since this set might contain many pseudogenes and misassembled transcripts, we decided to keep our filtered, but less complete set of genes.

Repetitive fraction in the *Aeluropus littoralis* Genome

In order to characterize the repetitive DNA fraction of *A. littoralis* the reads from the paired end WGS were used. Reads, comprising in total 0.26-fold genome coverage, were grouped based on sequence similarity into 33385 clusters containing from 2 to 21265 reads. Clusters included 32% of all reads, with the major 282 clusters representing at least 0.01% of the genome each. The clustering analysis revealed that 21.69% of the *A. littoralis* genome is composed of repetitive elements with nine satellite DNA families (satDNAs), nine transposable elements families (LTR-retrotransposons and LINE), two DNA transposons families (CACTA-like and Mutator-like), ribosomal DNA (35S and 5S) and microsatellites (**Table 2; Figure 4A**). The most abundant repetitive families were satDNAs, ~11% of the genome, with the five largest clusters being part of a superfamily named as AlSat140. Within this superfamily was possible to identified five variants: AlSat140a, AlSat140b, AlSat140c, AlSat70a and AlSat70b, with 85 to 96% similarity of the monomer sequence (**Figure 4B**). Beside these satDNAs, four other satDNAs families were identified in the genome: AlSat256, AlSat897, AlSat372 and AlSat80, with 0.62%, 0.42%, 0.03% and 0.02% of the genome, respectively. The LTR-like retrotransposons constituted 2.22% of the genome, with the Ty3/Gypsy superfamily exceeding 2.4 fold the genome proportion of the Ty1/Copia superfamily. Within the former, Tat/Retand, Tat/Ogre and Chromovirus were the only highly abundant lineages. Within Ty1/Copia retrotransposons were identified AleI, Ikeros and TAR lineage, with the last being more abundant. Microsatellites were identified in several different clusters comprising 3.23% of the genome (**Table 2; Figure 4A**).

Utility and Discussion

Based on our results, summarized in **Table 3**, the chromosome number of *A. littoralis* is 20 ($2n = 2x = 20$). We would like to draw attention to the point that in table 3 the, newly assembeled genome, which is still highly fragmented, is compared to almost fully assembled genomes. As pointed out before, the available information from prior published data [4, 6, 30] was contradicting concerning the presence of B chromosomes. The use of fluorescence *in situ* hybridization with a telomere-specific probe clearly indicated that metaphases where more than 20 chromatin units were counted are the result of extended NORs with only very thin chromatin fibres between the satellite and the corresponding chromosome. Such decondensed chromosomal rDNA sites were also described for *Lolium* and *Festuca* genotypes (Rocha et al. 2017). When only a DNA stain is applied distally located satellites can easily be miscounted as separate chromsomes. At least for the plant material analysed in the present paper the occurrence of B-chromosmes can be excluded. If the chromosome counts deposited in the Chromosome Counts Database [30], where the presence of B chromosomes in this species was reported, are indeed correct remains to be answered. However, also in the close related species *Aeluropus macrostachyus Hack*. the presence of B

chromosomes was described (*x* = 10 + 1B). In contrast, in *Aeluropus lagopoides (L.) Thwaites* (x = 10) no B chromosomes were detected [30]. Thus, the base chromosome number within the genus is stable with *x* = 10 A chromosomes.

As indicated in **Figure 4** approximately 85% of the genome information is covered by the presented sequencing approach. The BUSCO [34] analysis revealed ~63% of the proposed BUSCO genes. This relative low number might be underestimated due to the effect that BUSCO analysis are working very well in major species, where related species are well described, but can be more inaccurate in non-standard genomes like A. littoralis with difficult-to-assemble regions as stated by [35].

The number of annotated gene models with 15,916 is relative small, compared to other monocotyledonous plants (**Table S1**). However a substantial amount of genes is identified and the sequence information can be used for further research. In addition, the repeat fraction analysis revealed that 21.6% of the genome is composed by different repetitive elements, mainly by tandem repeat sequences distributed in several satellite DNA families (**Table 2**). The high abundance of the AlSat repeat family makes it likely that this provides a function as centromere building block. Small genomes are known to compromising low amounts of repetitive sequences, mostly less transposable elements, and being constituted mainly by tandem repeats, as satellite DNAs [36]. Thus, the repeat composition of *A. littoralis* is in agreement with this assumption of small genome composition.

In monocotyledon plants the plastid genome is maternally inherited and excluded from sexual recombination. Taking the highly conserved and overviewed chloroplast genome as proxy for the entire genome it can be stated that our sequencing approach covers preferentially genic regions, while repetitive sequences are not well assembled. The plastid genome also shows that no small genes (like t-RNA genes) are included in the annotation (**Table S2**).

As shown in **Figure 2**, *A. littoralis* is not only able to survive, but also to grow and develop on high soil salinity [17] and tolerate high salt concentrations in the soil up to 1 M sodium chloride. The plant is able to secrete salt via its salt glands leading to formation of salt crystals on the leaf surface [17, 37]. As reported before, we could also confirm the C_4 carbon fixation mechanism, based on Kranz anatomy (**Figure 2**). Based on ultrastructural analysis seems that the plant exhibits an unusual feature. The bundle sheath cells seem to be more compact, with an increase in stainable compounds. This might lead to a stronger differentiation of the tissues, and a better separation of the compartements, required for a more effective C_4 photosynthesis. Also the uptake of salt via the roots and formation of salt glands, followed by the secretion of salt is an interesting feature of this plant where the genome data might contribute to allow molecular insight into developmental and acclimation processes.

We are aware that our genomic study only is a glimpse into the genome of *A. littoralis* and can be complemented with a broader usage of biotechnological methods to reach a more comprehensive picture of this extraordinary species. However, here we wanted to show how versatile results it is possible to achieve using a simple WGS approach.

Methods

Plant material

Aeluropus littoralis seeds were collected from Isfahan province in Iran and plants were since cultivated at IPK Gatersleben (Germany) and Sari Agricultural Sciences and Natural Resources University (Iran). A specimen of the analyzed plants was deposited at the herbarium GAT under voucher number 70486. Sterilized seeds plated on full strength MS medium [38] with vitamins, 3% sucrose and 0.7% agar (pH 5.8). The cultures were incubated in germinator at 25 ± 2°C with 16 h light/8 h dark photoperiod at 100 µmol m^{-2} s^{-1} photon flux density using cool-white fluorescent light. Two weeks after germination, the seedlings were transferred to hydroponic culture containing Hoagland's solution [39]. Hoagland's nutrient solution comprised 3 mM KNO_3, 2 mM $Ca(NO_3)$, 2.1 mM $NH_4H_2PO_4$, 0.5 mM $MgSO_4$, 1 µM KCl, 25 µ M H_3BO_3, 2 µ M $MnSO_4$, 2 µ M $ZnSO_4$, 0.1 µ M $CuSO_4$, 0.1 µ M $(NH_4)_6Mo_7O_{24}$ and 20 µ M Fe(Na) EDTA, in demineralised H_2O buffered with 2 mM 2-(N-morpholino) ethanesulphonic acid, pH 5.5, set-out using KOH. Transferred plants were grown in a phytochamber at approximately 240 µmol $m^{-2}s^{-1}$ under photoperiodic conditions (16h light, 22°C /8h dark, 18°C) at relative humidity 70%.

For salt stress treatments soil grown plants (pots 12 cm diameter) were continuously watered with 1m NaCl 1 time per week.

Light and transmission electron microscopy

Aeluropus littoralis leaves of three biological replicates of plants grown under controlled conditions and exposed to salt stress were used for comparative histological and ultrastructural analysis. For this, cuttings of a size of 1x2 mm from the central part of fully developed leaves were used for combined conventional and microwave assisted chemical fixation, substitution and resin embedding as defined in the given protocol (**Supplemental Table 3**). Sectioning, histological staining, light and transmission electron microscopy analysis was performed as described [40].

Chromosome preparation and fluorescence *in situ* hybridization

Mitotic chromosomes were prepared from root tips pretreated in ice water for 24 h to accumulate synchronized cells at metaphase, fixed in Carnoy's fixative (ethanol and glacial acetic acid, 3:1 (v/v)) at room temperature for 20 h and kept in 70% ethanol at -20°C for later use. Fixed roots were digested in enzyme mixture (2% cellulose, 2% pectinase, 2% pectolyase in citrate buffer (0.01 M sodium citrate dihydrate and 0.01 M citric acid)) at 37°C for 30-40 min. Cell suspension from root meristems in the Carnoy's fixative was dropped onto slides on a hot plate at 50°C, slides were further fixed in the fixative for 1 min, air-dried and kept at 4°C.

The Arabidopsis-type telomere probe was labelled with fluorophore ATTO488 using nick translation labelling kit (Jena Bioscience). Fluorescence *in situ* hybridization was performed as described in [41] with pretreatment for 10 min in 45% acetic acid at room temperature, followed by 0.1% pepsine in 0.01N HCl at 37°C. Slides were applied with the hybridization mix (50% (v/v) formamide, 10% (w/v) dextran sulfate, 2× SSC, and 3 ng/µl of telomere probe) and denatured at 75°C for 2 min. After stringency wash in 2× SSC at 57°C for 20 min, chromosomes were counterstained with 4',6-diamidino-2-phenylindoline (DAPI). Images were captured using an epifluorescence microscope BX61 (Olympus) equipped with a cooled CCD camera (Orca ER, Hamamatsu) and pseudocolored with Adobe Photoshop.

Estimation of nuclear genome size

For estimation of nuclear genome size by flow cytometry, approximately 10 mm^2 of leaf tissue from individuals of *Aeluropus littoralis* populations was chopped with a sharp razor blade together with rougthly 5 mm^2 of leaf material of *Raphanus sativus* cv. 'Voran' (Genebank Gatersleben accession number: RA 34; 2C = 1.11 pg) as internal reference standard [32] in a Petri dish containing 1 ml Galbraith nuclei isolation buffer [42] supplemented with 1 % PVP- 25, 0.1 % Triton X-100, *DNase*-free *RNase* (50 µg/ml). The nuclei suspension was filtered through a 35-µm, mesh cell strainer cap to remove large fragments and stored on ice until measurement. The relative fluorescence intensities of around 7,000 to 10,000 events (nuclei) per sample were measured using a CyFLow Space flow cytometer (Sysmex-Partec, Germany) quipped with a 30 mW green solid state laser (532 nm). The absolute DNA amounts of samples were calculated based on the values of the G1 peak means.

Extraction of genomic DNA

DNA of *A. littoralis* was extracted according to Dellaporta procedures [43]. The quality and quantity of the extracted DNA were controlled by measuring absorbance at 260/280 nm using a NanoDrop spectrophotometer (Biochrom WPA Biowave II, UK). Further, the purity and integrity of DNA were tested by running on 0.7% agarose gel electrophoresis.

Illumina sequencing and sequence data pre-processing

Library preparation (Illumina TruSeq DNA Sample Prep Kit) and sequencing by synthesis using the Illumina HiSeq2500 device involved standard protocols from the manufacturer (Illumina, Inc., San Diego, CA, USA). The library was quantified by qPCR [44] and sequenced using the rapid run mode (on-board cluster generation, paired-end, 2 x 101 cycles. In total, 125,600,517 Illumina paired end reads were produced having a total output of residues of 42.5 Gb. Prior to the assembly process reads were quality trimmed using the clc_quality_trim module of CLC Genomics Workbench 11 with a minimal cut-off threshold of

Q30 and default settings on remaining parameters. 85.6% of reads and 83.77 % of residues passed this initial pre-processing. Subsequently, the quality of the sequence data checked using fastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). After this quality enrichment a genome coverage of 62-fold was reached.

De novo assembly construction

Our *A. littoralis de novo* sequence was constructed using CLC assembly cell version 4.3 and the quality trimmed WGS data. The *de novo* assembly pipeline was applied with automatic detection of best parameters by CLC assembly cell. In accordance with good practice, all contigs below a length threshold of 200 bp were removed. For purification of the constructed assembly we checked our constructed contigs for contamination by *E. coli* using BLAST+[45]. As parameter settings we used a sequence identity of 60% and a word size of 28. Critical contigs were fully removed if the BLASTN analysis resulted in a hit with length >500bp. For smaller contigs we reduced the minimal length of a hit 200 bp, while at the same time at least 10% of the length of the contig is identified as *E. coli* contamination. From the remaining sequences we removed contigs in case a bacterial origin was detected within the BLASTN analysis against the NCBI non-redundant nucleotide database nt. In addition, we filtered for contigs having a length of 500bp. The descriptive statistics of both datasets (200bp and 500bp) are described in **Table 1**. The list of all contigs is available at https://doi.ipk-gatersleben.de/DOI/ca99c593-ffdd-4d49-8eab-f1c891953776/d5b041b5-b2c1-4696-bc7c-bc5a32a0c7ec/2/1847940088.

Gene prediction and annotation

We used the purified WGS assembly without a threshold on contig sizes to predict gene models. Gene prediction was done with GeMoMa [46] using gene models of *Brachypodium distachyon* (Brachypodium_distachyon_v3.0, INSDC Assembly GCA_000005505.4, Feb 2018), *Oryza sativa* (IRGSP-1.0, INSDC Assembly GCA_001433935.1, Oct 2015) and *Sorghum bicolor* (Sorghum_bicolor_NCBIv3, INSDC Assembly GCA_000003195.3, Apr 2017) downloaded from Ensembl Plants [47]. In total, 15,916 gene models were predicted in 12,130 different contigs. For all detected gene models CDS (FASTA), protein sequence (FASTA) and genomic positions (GFF) are provided. We further investigated these dataset performing a gene annotation with AHRD version 2.0 (https://github.com/groupschoof/AHRD/) using UniProt, trembl and TAIR10 (downloaded January 4th 2016). For 13,921 genes (87.5 %) a functional annotation could be assigned. Complete data of gene models and functional annotation is available for download. The coding sequences of all annotated genes are available at https://doi.ipk-gatersleben.de/DOI/ac423f10-971e-481e-bcab-6ac261e27f5c/15d455e1-da91-4e78-82d8-7c7607cb05b9/2/1847940088 (provisional DOI). DOIs of datasets released in this manuscript were constructed using the e!DAL system [48].

Genome repeat fraction analysis

The repetitive fraction analysis was performed with 89 Mbp of reads of the total genomic DNA (0.26x genome coverage). Sequenced reads, after the quality trimmed above, were grouped with the graph-based clustering algorithm based on sequence similiraty, implemented in the RepeatExplorer pipeline [49]. The paired-end reads clustering was performed with a minimum overlap of 55% and a similarity of 90%. Three independent analyses were performed, using a different dataset of reads of the same sequencing, to confirm the proportions of each cluster within the total genome. Repeat annotation and classification was performed for those clusters with an abundance of at least >0.01%. For basic repeat classification, protein domains were identified using the tool ‘Find RT Domains’ within RepeatExplorer [49]. Searches for sequence similarity, using different databases (RepeatMasker and GenBank) were performed and graph layouts of individual clusters were examined using the SeqGrapheR program [49]. Satellite DNAs were identified based on the TAREAN tool implemented in the pipeline, graph layouts and further examined using DOTTER [50].

Abbreviations

BS bundle sheath

ESTs expressed sequence tag

FISH fluorescence *in situ* hybridization

Mbp Megabasepairs

M mesophyll

NOR nucleolus organizing region

PE paired end reads

satDNA satellite DNA

WGS whole-genome sequencing approach

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets generated during and/or analysed during the current study are available in the e!DAL repository:

https://doi.ipk-gatersleben.de/DOI/ac423f10-971e-481e-bcab-6ac261e27f5c/15d455e1-da91-4e78-82d8-7c7607cb05b9/2/1847940088

Competing interests

The authors declare that they have no competing interests

Funding

This research was supported by the Genetics and Agricultural Biotechnology Institute of Tabarestan (GABIT) and Leibniz Institute of Plant Genetics and Crop Plant Research (IPK Gatersleben).

Authors' contributions

SHH, MA, BD, YTK, AH, performed experiments; TS performed bioinformatic genome analysis, MAB performed repeat masker study; SHH, GN, SAMMM, TA, MK conceptualized work and wrote the manuscript.

Acknowledgements

We thank Ines Walde and Joerg Fuchs (Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany) for their excellent technical assistance on library preparation , Illumina sequencing and flow cytometric genome size measurements, repectively. Prof. Andrea Bräutigam for suggestions on genome assembly. Heike Müller for excellent photography.

References

1. Ben-Saad R, Ben-Ramdhan W, Zouari N, Azaza J, Mieulet D, Guiderdoni E, Ellouz R, Hassairi A: **Marker-free transgenic durum wheat cv. Karim expressing the AlSAP gene exhibits a high level of tolerance to salinity and dehydration stresses**. *Molecular breeding* 2012, **30**(1):521-533.
2. Watson L, Dallwitz MJ, Johnston CR: **Grass Genera of the World - 728 Detailed Descriptions from an Automated Database**. *Aust J Bot* 1986, **34**(2):223-230.
3. Peterson PM, Romaschenko K, Johnson G: **A classification of the Chloridoideae (Poaceae) based on multi-gene phylogenetic trees**. *Mol Phylogenet Evol* 2010, **55**(2):580-598.
4. Modarresi M, Nematzadeh GA, Moradian F, Alavi SM: **Identification and cloning of the Cu/Zn superoxide dismutase gene from halophyte plant Aeluropus littoralis**. *Genetika* 2012, **48**(1):130-134.
5. Zonneveld BJ, Leitch IJ, Bennett MD: **First nuclear DNA amounts in more than 300 angiosperms**. *Ann Bot* 2005, **96**(2):229-244.
6. Zouari N, Ben Saad R, Legavre T, Azaza J, Sabau X, Jaoua M, Masmoudi K, Hassairi A: **Identification and sequencing of ESTs from the halophyte grass Aeluropus littoralis**. *Gene* 2007, **404**(1-2):61-69.
7. Wang R: **Plant functional types and their ecological responses to salinization in saline grasslands, Northeastern China**. *Photosynthetica* 2004, **42**(2):511-519.
8. Barhoumi Z, Djebali W, Chaibi W, Abdelly C, Smaoui A: **Salt impact on photosynthesis and leaf ultrastructure of Aeluropus littoralis**. *J Plant Res* 2007, **120**(4):529-537.
9. Frey W, Kurschner H: **Photosynthetic Pathways and Ecological Distribution of Halophytes from Some Inland Salines of Turkey, Jordan and Iran**. *Flora* 1983, **173**(3-4):293-310.
10. Shomeril.A, Waisel Y: **Effect of Sodium-Chloride on Balance between C3-Carbon and C4-Carbon Fixation Pathways**. *Physiol Plantarum* 1973, **29**(2):190-193.
11. Bose J, Munns R, Shabala S, Gilliham M, Pogson B, Tyerman SD: **Chloroplast function and ion regulation in plants growing on saline soils: lessons from halophytes**. *J Exp Bot* 2017, **68**(12):3129-3143.
12. Brilhaus D, Brautigam A, Mettler-Altmann T, Winter K, Weber AP: **Reversible Burst of Transcriptional Changes during Induction of Crassulacean Acid Metabolism in Talinum triangulare**. *Plant Physiol* 2016, **170**(1):102-122.
13. Winter K, Holtum JA: **Facultative crassulacean acid metabolism (CAM) plants: powerful tools for unravelling the functional elements of CAM photosynthesis**. *J Exp Bot* 2014, **65**(13):3425-3441.
14. Saad RB, Romdhan WB, Zouari N, Azaza J, Mieulet D, Verdeil J-L, Guiderdoni E, Hassairi A: **Promoter of the AlSAP gene from the halophyte grass Aeluropus littoralis directs developmental-regulated, stress-inducible, and organ-specific gene expression in transgenic tobacco**. *Transgenic research* 2011, **20**(5):1003-1018.
15. Mesléard F, Ham LT, Boy V, van Wijck C, Grillas P: **Competition between an introduced and an indigenous species: the case of Paspalum paspalodes (Michx) Schribner and Aeluropus littoralis (Gouan) in the Camargue (southern France)**. *Oecologia* 1993, **94**(2):204-209.
16. Barhoumi Z, Djebali W, Chaïbi W, Abdelly C, Smaoui A: **Salt impact on photosynthesis and leaf ultrastructure of Aeluropus littoralis**. *Journal of plant research* 2007, **120**(4):529-537.
17. Barhoumi Z, Djebali W, Smaoui A, Chaibi W, Abdelly C: **Contribution of NaCl excretion to salt resistance of Aeluropus littoralis (Willd) Parl**. *J Plant Physiol* 2007, **164**(7):842-850.
18. Gulzar S, Khan MA, Ungar IA: **Effects of salinity on growth, ionic content, and plant-water status of Aeluropus lagopoides**. *Commun Soil Sci Plan* 2003, **34**(11-12):1657-1668.
19. Gulzar S, Khan MA: **Seed germination of a halophytic grass Aeluropus lagopoides**. *Ann Bot-London* 2001, **87**(3):319-324.
20. Azri W, Barhoumi Z, Chibani F, Borji M, Bessrour M, Mliki A: **Proteomic responses in shoots of the facultative halophyte Aeluropus littoralis (Poaceae) under NaCl salt stress**. *Funct Plant Biol* 2016, **43**(11):1028-1047.
21. Saad RB, Zouari N, Ramdhan WB, Azaza J, Meynard D, Guiderdoni E, Hassairi A: **Improved drought and salt stress tolerance in transgenic tobacco overexpressing a novel A20/AN1 zinc-finger "AlSAP" gene isolated from the halophyte**

grass Aeluropus littoralis. *Plant molecular biology* 2010, **72**(1-2):171-190.

22. Rezvani M, Zaefarian F, Miransari M, Nematzadeh GA: **Uptake and translocation of cadmium and nutrients by Aeluropus littoralis**. *Archives of Agronomy and Soil Science* 2012, **58**(12):1413-1425.
23. Barhoumi Z, Djebali W, Abdelly C, Chaïbi W, Smaoui A: **Ultrastructure of Aeluropus littoralis leaf salt glands under NaCl stress**. *Protoplasma* 2008, **233**(3-4):195-202.
24. Ben Romdhane W, Ben-Saad R, Meynard D, Verdeil JL, Azaza J, Zouari N, Fki L, Guiderdoni E, Al-Doss A, Hassairi A: **Ectopic Expression of Aeluropus littoralis Plasma Membrane Protein Gene AlTMP1 Confers Abiotic Stress Tolerance in Transgenic Tobacco by Improving Water Status and Cation Homeostasis**. *Int J Mol Sci* 2017, **18**(4).
25. Ben Saad R, Fabre D, Mieulet D, Meynard D, Dingkuhn M, Al-Doss A, Guiderdoni E, Hassairi A: **Expression of the Aeluropus littoralis AlSAP gene in rice confers broad tolerance to abiotic stresses through maintenance of photosynthesis**. *Plant Cell Environ* 2012, **35**(3):626-643.
26. Ben Saad R, Zouari N, Ben Ramdhan W, Azaza J, Meynard D, Guiderdoni E, Hassairi A: **Improved drought and salt stress tolerance in transgenic tobacco overexpressing a novel A20/AN1 zinc-finger "AlSAP" gene isolated from the halophyte grass Aeluropus littoralis**. *Plant Mol Biol* 2010, **72**(1-2):171-190.
27. Ghneim-Herrera T, Selvaraj MG, Meynard D, Fabre D, Pena A, Ben Romdhane W, Ben Saad R, Ogawa S, Rebolledo MC, Ishitani M *et al*: **Expression of the Aeluropus littoralis AlSAP Gene Enhances Rice Yield under Field Drought at the Reproductive Stage**. *Front Plant Sci* 2017, **8**:994.
28. Liu J, Zhang S, Dong L, Chu J: **Incorporation of Na+/H+ antiporter gene from Aeluropus littoralis confers salt tolerance in soybean (Glycine max L**. *Indian J Biochem Biophys* 2014, **51**(1):58-65.
29. Liu Q, Zhao NX, Hao G: **Inflorescence structures and evolution in subfamily Chloridoideae (Gramineae)**. *Plant Syst Evol* 2005, **251**(2-4):183-198.
30. Rice A, Glick L, Abadi S, Einhorn M, Kopelman NM, Salman-Minkov A, Mayzel J, Chay O, Mayrose I: **The Chromosome Counts Database (CCDB) - a community resource of plant chromosome numbers**. *New Phytol* 2015, **206**(1):19-26.
31. Richardson KC, Jarett L, Finke EH: **Embedding in epoxy resins for ultrathin sectioning in electron microscopy**. *Stain Technol* 1960, **35**:313-323.
32. Schmidt-Lebuhn A, Fuchs J, Kessler M: **Flow cytometric measurements do not reveal different ploidy levels in Minthostachys (Lamiaceae)**. *Plant Systematics and Evolution* 2008, **271**(1-2):123-128.
33. Stanke M, Schoffmann O, Morgenstern B, Waack S: **Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources**. *BMC Bioinformatics* 2006, **7**:62.
34. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs**. *Bioinformatics* 2015, **31**(19):3210-3212.
35. Rhie A, Walenz BP, Koren S, Phillippy AM: **Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies**. *Genome Biol* 2020, **21**(1):245.
36. Macas J, Novak P, Pellicer J, Cizkova J, Koblizkova A, Neumann P, Fukova I, Dolezel J, Kelly LJ, Leitch IJ: **In Depth Characterization of Repetitive DNA in 23 Plant Genomes Reveals Sources of Genome Size Variation in the Legume Tribe Fabeae**. *PLoS One* 2015, **10**(11):e0143424.
37. Barhoumi Z, Djebali W, Abdelly C, Chaibi W, Smaoui A: **Ultrastructure of Aeluropus littoralis leaf salt glands under NaCl stress**. *Protoplasma* 2008, **233**(3-4):195-202.
38. Murashige T, Skoog F: **A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures**. *Physiol Plantarum* 1962, **15**(3):473-497.
39. Arnon DI, Hoagland DR: **A Comparison of Water Culture and Soil as Media for Crop Production**. *Science* 1939, **89**(2318):512-514.
40. Daghma DS, Kumlehn J, Melzer M: **The use of cyanobacteria as filler in nitrocellulose capillaries improves ultrastructural preservation of immature barley pollen upon high pressure freezing**. *J Microsc* 2011, **244**(1):79-84.

41. Aliyeva-Schnorr L, Ma L, Houben A: **A Fast Air-dry Dropping Chromosome Preparation Method Suitable for FISH in Plants**. *J Vis Exp* 2015(106):e53470.
42. Galbraith DW, Harkins KR, Maddox JM, Ayres NM, Sharma DP, Firoozabady E: **Rapid flow cytometric analysis of the cell cycle in intact plant tissues**. *Science* 1983, **220**(4601):1049-1051.
43. Dellaporta SL, Wood J, Hicks JB: **A plant DNA minipreparation: version II**. *Plant molecular biology reporter* 1983, **1**(4):19-21.
44. Mascher M, Richmond TA, Gerhardt DJ, Himmelbach A, Clissold L, Sampath D, Ayling S, Steuernagel B, Pfeifer M, D'Ascenzo M *et al*: **Barley whole exome capture: a tool for genomic research in the genus Hordeum and beyond**. *Plant J* 2013, **76**(3):494-505.
45. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: **BLAST+: architecture and applications**. *BMC Bioinformatics* 2009, **10**:421.
46. Keilwagen J, Wenk M, Erickson JL, Schattat MH, Grau J, Hartung F: **Using intron position conservation for homology-based gene prediction**. *Nucleic Acids Res* 2016, **44**(9):e89.
47. Bolser DM, Staines DM, Perry E, Kersey PJ: **Ensembl Plants: Integrating Tools for Visualizing, Mining, and Analyzing Plant Genomic Data**. *Methods Mol Biol* 2017, **1533**:1-31.
48. Arend D, Lange M, Chen J, Colmsee C, Flemming S, Hecht D, Scholz U: **e!DAL–a framework to store, share and publish research data**. *BMC Bioinformatics* 2014, **15**:214.
49. Novak P, Neumann P, Pech J, Steinhaisl J, Macas J: **RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads**. *Bioinformatics* 2013, **29**(6):792-793.
50. Sonnhammer EL, Durbin R: **A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis**. *Gene* 1995, **167**(1-2):GC1-10.

Tables

Table 1 Descriptive statistics of WGS assembly

parameter	A. littoralis WGS assembly	A. littoralis WGS assembly (>500bp)
number of contigs	182,747	113,845
total number of bases	300,381,201	277,993,896
minimal contigs length	200	500
maximal contigs length	69,774	69,774
N25 contig length (bp)	7,574	8,048
N50 contig length (bp)	3,649	4,074
N75 contig length (bp)	1,477	1,869
GC content	0.44	0.44

Table 2 Genome proportion of the repetitive elements identified in A. littoralis.

	Repetitive element		Total genome %
	Satellite	AlSat140a	3.98
		AlSat70a	2.29
		AlSat140b	1.60
		AlSat140c	1.33
		AlSat256	0.71
		AlSat897	0.48
		AlSat71	0.80
		AlSat372	0.03
		AlSat80	0.02
	Microsatellites		3.23
Class I			
	LTR-Ty3/Gypsy	Tat/Ogre	0.62
		Chromo/Tekay	0.32
		Chromo/Reina	0.01
		Chromo/CRM	0.02
		Tat/Retand	0.60
	LTR-Ty1/Copia	AleI	0.17
		Ikeros	0.18
		TAR	0.30
	LINE		0.04
Class II			
	DNA Transposons	CACTA	0.06
		Mutator	0.01
	rDNA		1.04
	Unclassified		3.84
	Total		21.69

Table 3 Statistics of Aleuropus genome assembly

		Oryza sativa	*Sorghum bicolor*	*Brachypodium distachyon*	*Aeluropus littoralis*
	Assembly	*IRGSP-1.0, INSDC*	*Sorghum_bicolor_NCBIv3*	Brachypodium_distachyon_v3.0	*this study*
Chromosome number	(2n)	24	20	10	20
Genome	Size	500 Mb	730 Mb	355 Mb	354 Mb
Sequencing	covered	375,049,285	675,363,888	270,739,641	300,381,201
Sequenced	%	75	93	76	85
Gene number	annot.	37849	34118	34310	15916

Figures

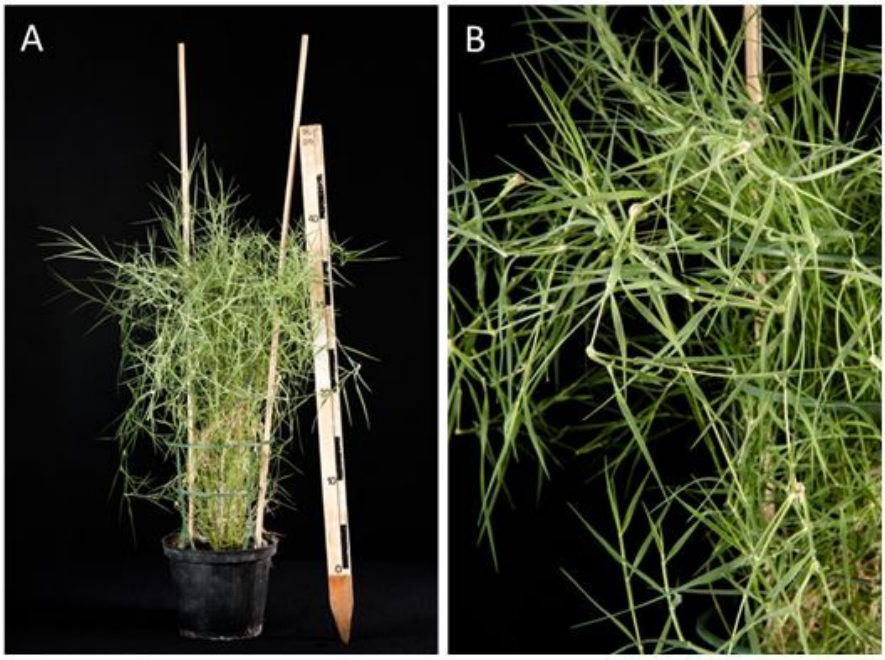

Figure 1

A Aeluropus littoralis cultivated in pots under greenhouse condition, B Enlargement of leaves. Black and white box on the ruler represents 5 cm.

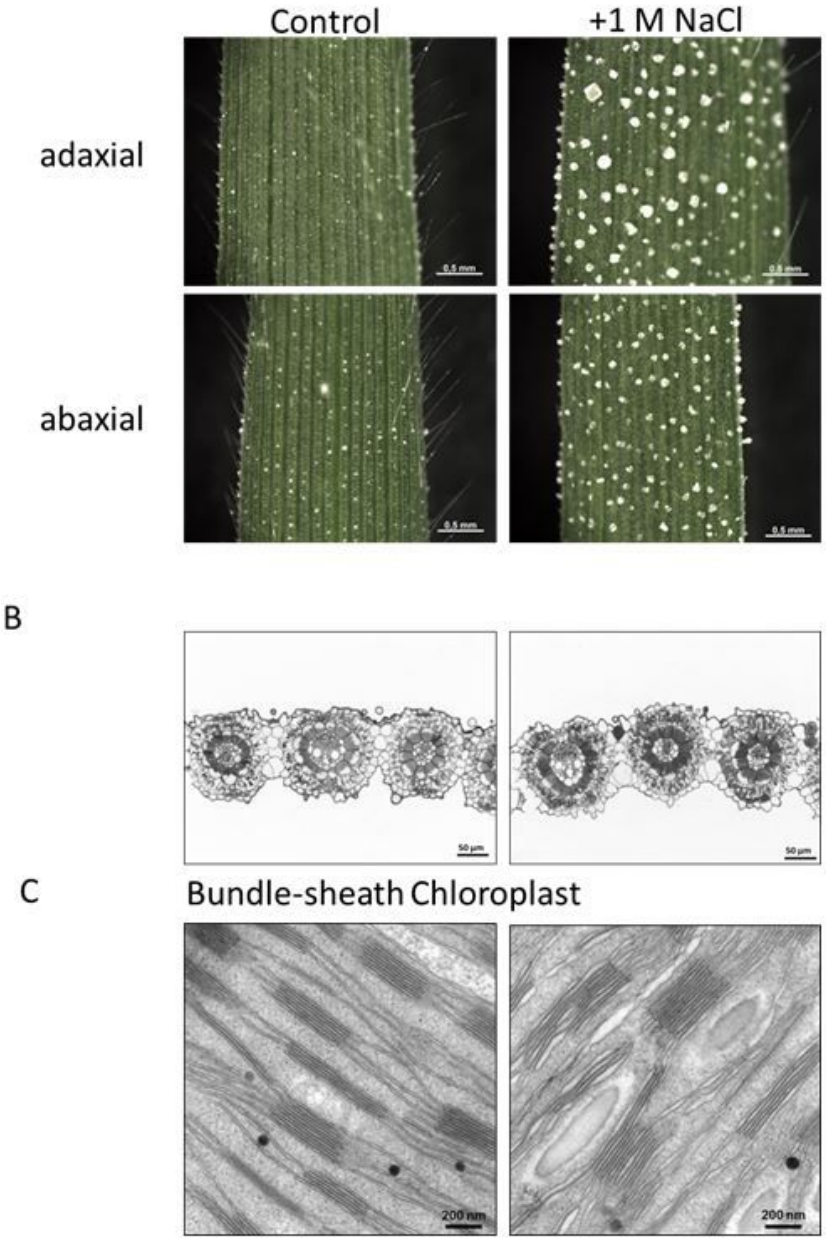

Figure 2

A Leaf surface of young leaved developed in pots under greenhouse conditions at different watering regiemes. left: control (tap water) and right: salt watering (1 M NaCl). Shown are adaxial and abaxial sides of the leaves including formed salt crystals at salt watering condition. B. Microscopic section of leaves after methyblue/azur II stain showing Kranz anatomy, typical indication for C4 type plant. Bundle sheath cells appear darker at salt treatment. C. SEM analysis of thylakoid structure in bundle sheath cell chloroplasts. Grana staples are larger and show spaces between layers at salt treatment.

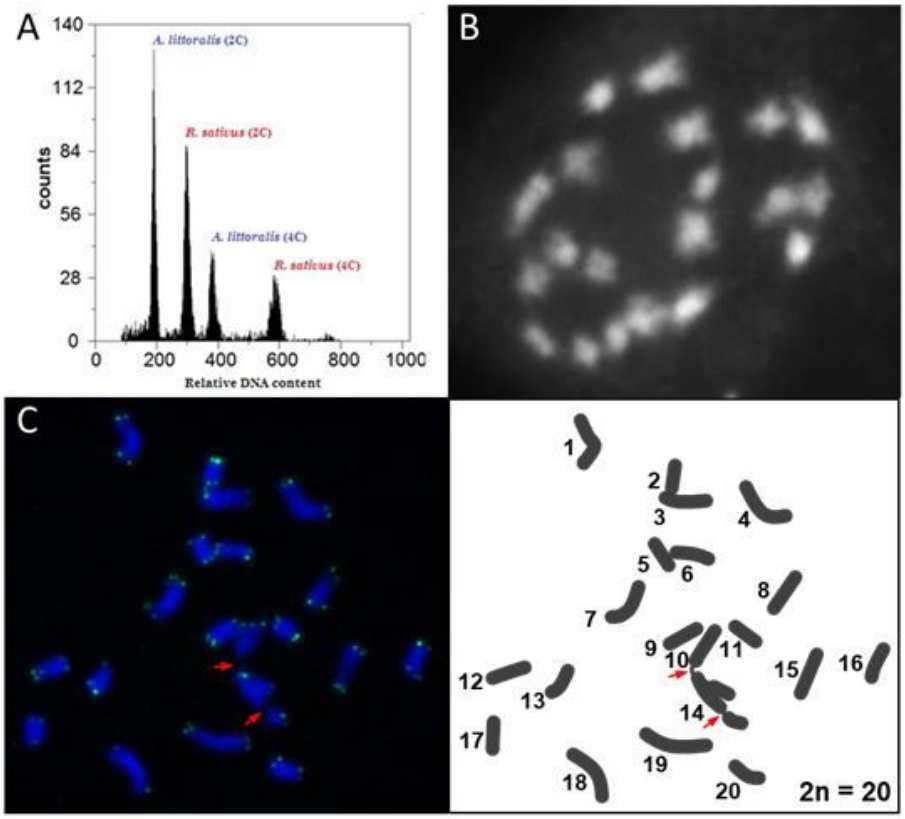

Figure 3

A Genome size estimation and chromosome counting of Aeluropus littoralis. Measurements of Aeluropus littoralis genome size by flow cytometry. Based on the 2C-value of 1.11 pg for Raphanus sativus, the average DNA content of diploid Aeluropus littoralis was found to be 0.724 ± 0.01 pg/2C (354 Mbp/1C). Representative flow cytometric histogram representing the relative fluorescence intensity of leaf nuclei in comparison to the internal reference standard of Pisum sativum. B DAPI stained metaphase chromosomes isolated from root tip meristem. C FisH of telomereic structures indicate the absence of B chromosomes

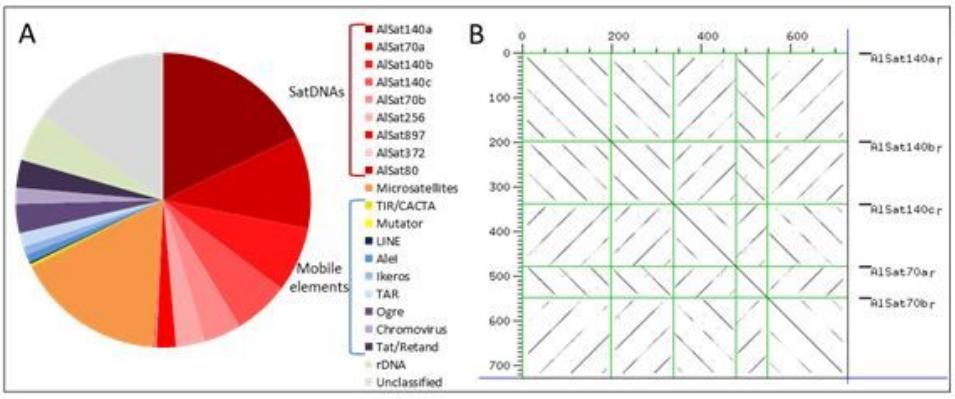

Figure 4

A Aeluropus littoralis genome repetitive composition, satDNAs are the most abundant repetitive element within the genome. B Dot-plot showing the sequences similarity between the different variants of the same satellite family, AlSat140.

Supplementary Files

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

- FigS1.JPG
- TableS1.xlsx
- TableS2plastid.xlsx
- Stab3Lightandtransmissionelectronmicroscopy.docx