

DATA NOTE

Open Access



BeetRepeats: reference sequences for genome and polymorphism annotation in sugar beet and wild relatives

Nicola Schmidt¹ , Sophie Maiwald¹ , Ludwig Mann¹ , Beatrice Weber¹ , Kathrin M. Seibt¹ , Sarah Breitenbach¹, Susan Liedtke¹, Gerhard Menzel¹ , Bernd Weisshaar² , Daniela Holtgräwe² and Tony Heitkam^{1,3*}

Abstract

Objectives Despite the advances in genomics, repetitive DNAs (repeats) are still difficult to sequence, assemble, and identify. This is due to their high abundance and diversity, with many repeat families being unique to the organisms in which they were described. In sugar beet, repeats make up a significant portion of the genome (at least 53%), with many repeats being restricted to the beet genera, *Beta* and *Patellifolia*. Over the course of over 30 years and many repeat-based studies, over a thousand reference repeat sequences for beet genomes have been identified and many experimentally characterized (e.g. physically located on the chromosomes). Here, we present the collection of these reference repeat sequences for beets.

Data description The BeetRepeats_v1.0 resource is a comprehensive compilation of all characterized repeat families, including satellite DNAs, ribosomal DNAs, transposable elements and endogenous viruses. The genomes covered are those of sugar beet and closely related wild beets (genera *Beta* and *Patellifolia*) as well as *Chenopodium quinoa* and *Spinacia oleracea* (all belonging to the Amaranthaceae). The reference sequences are in fasta format and comprise well-characterized repeats from both repeat categories (dispersed/mobile as well as tandemly arranged). The database is suitable for the RepeatMasker and RepeatExplorer2 pipelines and can be used directly for any repeat annotation and repeat polymorphism detection purposes.

Keywords Sugar beet, *Beta vulgaris*, *Patellifolia*, Repetitive DNA, Transposable elements, Satellite DNAs, Genome annotation

Objective

Due to its roles in beet evolution and variability, sugar beet's (*Beta vulgaris* subsp. *vulgaris*) repeatome has been a subject of interest for over 30 years. Starting with the detection of repeat-derived ladder patterns using Southern hybridization experiments [1], more detailed studies became possible, including wet lab (i.e. fluorescent in situ hybridization; e.g. [2, 3]) as well as bioinformatics methods (i.e. read clustering [4–6]).

A significant portion (at least 53% [5]) of the sugar beet genome consists of repeats, comprising a great number of different transposable elements (TEs) as well as

*Correspondence:

Tony Heitkam

tony.heitkam@tu-dresden.de; tony.heitkam@rwth-aachen.de

¹ Faculty of Biology, Technische Universität Dresden, 01069 Dresden, Germany

² Fakultät of Biology & CeBiTec, Universität Bielefeld, 33615 Bielefeld, Germany

³ Institute of Biology I, RWTH Aachen University, 52056 Aachen, Germany



tandem repeats. Compiling and unifying data from over 50 publications and theses, we here provide a downloadable and easy-to-use resource of the repeat profiles in beet genomes [7].

With this data note, we provide a comprehensive collection of all characterized repeat families in sugar beet and closely related wild beet genomes [7], facilitating genome annotation as well as the investigation of evolutionary trajectories of TEs and their host genomes.

Data description

We have collected repetitive DNA sequences that are representative for all major repeat families in genomes of the crop sugar beet (*Beta vulgaris* subsp. *vulgaris*) and related wild beets (genera *Beta* and *Patellifolia*), and provide them in fasta format [7]. Furthermore, we added repeats identified in two further Amaranthaceae species (*Chenopodium quinoa* and *Spinacia oleracea*). In detail, the beet repeatomes are represented by:

- 223 non-long terminal repeat (non-LTR) retrotransposons, with 100 long interspersed nuclear elements (Belline LINES) [8, 9] and 123 short interspersed nuclear elements (AmaSINES) [10];
- 61 non-autonomous LTR retrotransposons, with 60 terminal repeat retrotransposons in miniature (TRIMs) [7, 11] and one large retrotransposon derivative (LARD) [7];
- 355 Ty1-copia retrotransposons, with 220 Retrofit sequences, 11 Oryco/Ivana sequences, 87 Tork sequences, 9 SIRE sequences, and 28 Bianca sequences [7, 12, 13];
- 69 Ty3-gypsy retrotransposons, with 25 chromoviruses [3, 14], 27 errantiviruses/Athila [15], and 17 Tat sequences [7];
- 3 endogenous pararetroviruses (beetEPRVs) [16];
- 299 DNA transposons, with 12 EnSpm/CACTA sequences [17], 3 autonomous and 116 non-autonomous hAT sequences (BvhAT and BvhATpin MITEs) [18], 51 autonomous and 90 non-autonomous PIF/Harbinger sequences (BvPIF/Pong and BvPIF/Pong

- MITEs) [7], one autonomous and 24 non-autonomous Tc1_Mariner sequences (Vulmar and VulMITEs) [17, 19], and 2 helitron sequences [7];
- 82 satellite DNAs [2, 5, 20–31];
- 21 minisatellites/tandem repeats [7, 27, 32, 33];
- 3 rDNAs (two variants of the 5S rDNA and one 45S rDNA sequence) [34, 35].

This list is further detailed in Data file 1 (see Table 1).

The BeetRepeats fasta resource contains in silico consensus sequences as well as exemplary, representative copies. It is formatted to meet the requirements for a ‘custom repeat database’ utilized by the RepeatExplorer2 pipeline [36]. Due to the absence of a respective category in the RepeatExplorer2 annotation, all tandem repeats listed in our resource (except rDNA) were classified as satellite DNAs [7].

As there are several assemblies of the sugar beet genome available, we provide an annotation of the repeats from our database within the three different sugar beet assemblies EL10 [37], 2320BvONT_v1.0 [38], and RefBeet1.5 (<https://jbrowse.cebitec.uni-bielefeld.de/RefBeet1.5/>) as GFF files (see Table 1). To create these annotation files, we used the RepeatMasker pipeline [39] with standard parameters (performing softmasking instead of hardmasking and deactivating the low complexity masking).

Limitations

- Whereas the repeatome of sugar beet should be completely covered with this database, repeat identification and characterization in wild beet genomes are still ongoing. Thus, the database is under constant expansion and requires regular updates.
- Due to different requirements for custom databases, it may be necessary to reformat our database in order to use it with certain software programs. For instance, to acquire a detailed summary table by RepeatMasker, additional information on repeat classification in the sequence names have to be

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File type (file extension)	Data repository and identifier (DOI or accession number)
Data set 1	BeetRepeatDB_v1.0.fasta	Fasta file (.fa)	Zenodo (https://doi.org/10.5281/zenodo.8255813) [7]
Data set 2	BeetRepeatDB_v1.0_at_EL10.gff	GFF file (.gff)	Zenodo (https://doi.org/10.5281/zenodo.8255813) [7]
Data set 3	BeetRepeatDB_v1.0_at_2320BvONT_v1.0.gff	GFF file (.gff)	Zenodo (https://doi.org/10.5281/zenodo.8255813) [7]
Data set 4	BeetRepeatDB_v1.0_at_RefBeet1.5.gff	GFF file (.gff)	Zenodo (https://doi.org/10.5281/zenodo.8255813) [7]
Data file 1	BeetRepeatDB_v1.0-Content.docx	Microsoft Word Document (.docx)	Zenodo (https://doi.org/10.5281/zenodo.8255813) [7]

shortened to only indicate the class and subclass of the respective repeat.

- Diverged repeated sequences (e.g. old and/or lowly abundant variants with an accumulation of mutations) are not completely covered by our representative sequences. Therefore, it is to be expected that repeat masking of beet genomes using our database results in lower repetitive genome proportions than the actual repeat fraction.
- Since we focused largely on beets, we included only few selected repeats from *C. quinoa* and *S. oleracea* in our resource [7]. The repeats from these more distantly related plants were only included, if they represented derivatives of the listed beet repeats.

Abbreviations

TE	Transposable element
LTR	Long terminal repeat
LINE	Long interspersed nuclear element
SINE	Short interspersed nuclear element
TRIM	Terminal repeat retrotransposon in miniature
LARD	Large retrotransposon derivative
EPRV	Endogenous pararetrovirus
MITE	Miniature inverted-repeat transposable element
rDNA	Ribosomal DNA

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13104-024-06993-4>.

Additional file 1.

Acknowledgements

Over time, many scientists have worked on the repeat landscape of cultivated and wild beets. Especially the group around late Prof. Thomas Schmidt has accumulated a large knowledge around this topic; we counted over 30 student theses on beet repeats over the years! Regarding authorship, we included those who have been instrumental in compiling the database at the current time point. Here, we thank the people and former team members who contributed in shedding light onto individual repeats: In alphabetical order, these are Susanne Antoniotti, Markus Badstübner, Hans-Ulrich Balcke, Ekaterina Bannack, Juliane Bettig, Daryna Decheyeva, Christine Desel, Janine Epperlein, Conny Fiege, Frank Gindullis, Annelie Gutsch, Hong Bich Ha, Michel Heidecker, Axel Horn, Cordula John, Luise Keßler, Jessica Klekar, Katharina Kölling, Teresa Kowar, Sybille Kubis, Katrin Lenz, Anna Voigt, Ines Walter, Christina Wäsch, Torsten Wenke, Cora Wollrab, Falk Zakrzewski. We also thank our cooperators in beet repeats projects, namely Britta Schulz (KWS SAAT SE & Co. KGaA) as well as Pat Heslop-Harrison and Trude Schwarzbacher (University of Leicester), who were among the first to believe in the importance of repetitive DNA sequences in plant genomes. We also thank Heinz Himmelbauer, André Minoche and Juliane Dohm, who have led the first beet reference genome sequence about a decade ago. Similarly, Mitch McGrath is acknowledged for being one of the first to openly share many of the sequences that he had produced in the early years of genome sequencing – allowing the group unprecedented access to many beet repeat sequences.

Author contributions

NS, SM, BW, KMS, GM and TH collected the sequences. NS, SM, LM, SB, SL, BW, DH, and TH wrote the manuscript. All authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. This work was supported by the German Federal Ministry of Education and Research (call „Epigenetics: Opportunities for Plant Research“, grant 031B1221).

Availability of data and materials

The data described in this Data note can be freely and openly accessed on Zenodo under <https://doi.org/https://doi.org/10.5281/zenodo.8255813> [7]. Please see Table 1 for details and links to the data.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 6 June 2024 Accepted: 1 November 2024

Published online: 27 November 2024

References

- Schmidt T, Junghans H, Metzlaß M. Construction of *Beta procumbens*-specific DNA probes and their application for the screening of *B. vulgaris* × *B. procumbens* (2n = 19) addition lines. *Theoret Appl Genet*. 1990;79:177–81.
- Gindullis F, Desel C, Galasso I, Schmidt T. The large-scale organization of the centromeric region in *Beta* species. *Genome Res*. 2001;11:253–65.
- Weber B, Heitkam T, Holtgräwe D, Weisshaar B, Minoche AE, Dohm JC, et al. Highly diverse chromoviruses of *Beta vulgaris* are classified by chromodomains and chromosomal integration. *Mob DNA*. 2013;4:1–16.
- Kowar T, Zakrzewski F, Macas J, Kobližková A, Viehoveer P, Weisshaar B, et al. Repeat composition of CenH3-chromatin and H3K9me2-marked heterochromatin in sugar beet (*Beta vulgaris*). *BMC Plant Biol*. 2016;16:120.
- Schmidt N, Sielemann K, Breitenbach S, Fuchs J, Pucker B, Weisshaar B, et al. Repeat turnover meets stable chromosomes: repetitive DNA sequences mark speciation and gene pool boundaries in sugar beet and wild beets. *Plant J*. 2023. <https://doi.org/10.1111/tpj.16599>.
- Mann L, Balasch K, Schmidt N, Heitkam T. High-fidelity (repeat) consensus sequences from short reads using combined read clustering and assembly. *BMC Genom*. 2024;25:109.
- Schmidt N, Maiwald S, Mann L, Weber B, Seibt KM, Breitenbach S, et al. BeetRepeats: reference sequences for genome and polymorphism annotation in sugar beet and wild relatives. 2024. Zenodo. <https://doi.org/10.5281/zenodo.8255813>.
- Heitkam T, Schmidt T. BNR - a LINE family from *Beta vulgaris* contains an RRM domain in open reading frame 1 and defines a L1 subclade present in diverse plant genomes. *Plant J*. 2009;59:872–82.
- Heitkam T, Holtgräwe D, Dohm JC, Minoche AE, Himmelbauer H, Weisshaar B, et al. Profiling of extensively diversified plant LINEs reveals distinct plant-specific subclades. *Plant J*. 2014;79:385–97.
- Schwichtenberg K, Wenke T, Zakrzewski F, Seibt KM, Minoche A, Dohm JC, et al. Diversification, evolution and methylation of short interspersed nuclear element families in sugar beet and related Amaranthaceae species. *Plant J*. 2016;85:229–44.
- Maiwald S, Weber B, Seibt KM, Schmidt T, Heitkam T. The Cassandra retrotransposon landscape in sugar beet (*Beta vulgaris*) and related Amaranthaceae: recombination and re-shuffling lead to a high structural variability. *Ann Bot*. 2021;127:91–109.
- Weber B, Wenke T, Frömmel U, Schmidt T, Heitkam T. The Ty1-copia families SALIRE and Cotzilla populating the *Beta vulgaris* genome show remarkable differences in abundance, chromosomal distribution, and age. *Chromosome Res*. 2010;18:247–63.

13. Brandes A, Heslop-Harrison JS, Kamm A, Kubis S, Doudrick RL, Schmidt T. Comparative analysis of the chromosomal and genomic organization of Ty1-copia-like retrotransposons in pteridophytes, gymnosperms and angiosperms. *Plant Mol Biol*. 1997;33:11–21.
14. Weber B, Schmidt T. Nested Ty3-gypsy retrotransposons of a single *Beta procumbens* centromere contain a putative chromodomain. *Chromosome Res*. 2009;17:379–96.
15. Wollrab C, Heitkam T, Holtgräwe D, Weisshaar B, Minoche AE, Dohm JC, et al. Evolutionary reshuffling in the Errantivirus lineage Elbe within the *Beta vulgaris* genome. *Plant J*. 2012;72:636–51.
16. Schmidt N, Seibt KM, Weber B, Schwarzacher T, Schmidt T, Heitkam T. Broken, silent, and in hiding: tamed endogenous pararetroviruses escape elimination from the genome of sugar beet (*Beta vulgaris*). *Ann Bot*. 2021;128:281–99.
17. Jacobs G, Dechyeva D, Menzel G, Dombrowski C, Schmidt T. Molecular characterization of Vulmar1, a complete mariner transposon of sugar beet and diversity of mariner- and En/Spm-like sequences in the genus *Beta*. *Genome*. 2004;47:1192–201.
18. Menzel G, Krebs C, Diez M, Holtgräwe D, Weisshaar B, Minoche AE, et al. Survey of sugar beet (*Beta vulgaris* L.) hAT transposons and MITE-like hATpin derivatives. *Plant Mol Biol*. 2012;78:393–405.
19. Menzel G, Dechyeva D, Keller H, Lange C, Himmelbauer H, Schmidt T. Mobilization and evolutionary history of miniature inverted-repeat transposable elements (MITEs) in *Beta vulgaris* L. *Chromosome Res*. 2006;14:831–44.
20. Schmidt T, Jung C, Metzlaß M. Distribution and evolution of two satellite DNAs in the genus *Beta*. *Theoret Appl Genetics*. 1991;82:793–9.
21. Schmidt T, Heslop-Harrison JS. Variability and evolution of highly repeated DNA sequences in the genus *Beta*. *Genome*. 1993;36:1074–9.
22. Schmidt T, Heslop-Harrison JS. High-resolution mapping of repetitive DNA by *in situ* hybridization: molecular and chromosomal features of prominent dispersed and discretely localized DNA families from the wild beet species *Beta procumbens*. *Plant Mol Biol*. 1996;30:1099–113.
23. Kubis S, Heslop-Harrison J, Schmidt T. A family of differentially amplified repetitive DNA sequences in the genus *Beta* reveals genetic variation in *Beta vulgaris* subspecies and cultivars. *J Mol Evol*. 1997;44:310–20.
24. Gao D, Schmidt T, Jung C. Molecular characterization and chromosomal distribution of species-specific repetitive DNA sequences from *Beta corolliflora*, a wild relative of sugar beet. *Genome*. 2000;43:1073–80.
25. Dechyeva D, Gindullis F, Schmidt T. Divergence of satellite DNA and interspersed of dispersed repeats in the genome of the wild beet *Beta procumbens*. *Chromosome Res*. 2003;11:3–21.
26. Dechyeva D, Schmidt T. Molecular organization of terminal repetitive DNA in *Beta* species. *Chromosome Res*. 2006;14:881–97.
27. Zakrzewski F, Wenke T, Holtgräwe D, Weisshaar B, Schmidt T. Analysis of a c0t-1 library enables the targeted identification of minisatellite and satellite families in *Beta vulgaris*. *BMC Plant Biol*. 2010;10:1–14.
28. Zakrzewski F, Weisshaar B, Fuchs J, Bannack E, Minoche AE, Dohm JC, et al. Epigenetic profiling of heterochromatic satellite DNA. *Chromosoma*. 2011;120:409–22.
29. Zakrzewski F, Weber B, Schmidt T. A molecular cytogenetic analysis of the structure, evolution, and epigenetic modifications of major DNA sequences in centromeres of *Beta* species. In: Jiang J, Birchler JA, editors. *Plant centromere biology*. Oxford: John Wiley & Sons; 2013. p. 39–55.
30. Ha BH. Structure, organization, and evolution of satellite DNAs in species of the genera *Beta* and *Patellifolia*. Doctoral dissertation, Technische Universität Dresden. 2018. <https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa-238083>.
31. Heitkam T, Weber B, Walter I, Liedtke S, Ost C, Schmidt T. Satellite DNA landscapes after allotetraploidization of quinoa (*Chenopodium quinoa*) reveal unique A and B subgenomes. *Plant J*. 2020;103:32–52.
32. Zakrzewski F, Schubert V, Viehoveer P, Minoche AE, Dohm JC, Himmelbauer H, et al. The CHH motif in sugar beet satellite DNA: a modulator for cytosine methylation. *Plant J*. 2014;78:937–50.
33. Li N, Li X, Zhou J, Yu L, Li S, Zhang Y, et al. Genome-wide analysis of transposable elements and satellite DNAs in *Spinacia* species to shed light on their roles in sex chromosome evolution. *Front Plant Sci*. 2021;11: 575462.
34. Schmidt T, Schwarzacher T, Heslop-Harrison JS. Physical mapping of rRNA genes by fluorescent *in-situ* hybridization and structural analysis of 5S rRNA genes and intergenic spacer sequences in sugar beet (*Beta vulgaris*). *Theoret Appl Genetics*. 1994;88:629–36.
35. Paesold S, Borchardt D, Schmidt T, Dechyeva D. A sugar beet (*Beta vulgaris* L.) reference FISH karyotype for chromosome and chromosome-arm identification, integration of genetic linkage groups and analysis of major repeat family distribution. *Plant J*. 2012;72:600–11.
36. Novák P, Neumann P, Macas J. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nat Protoc*. 2020;15:3745–76.
37. McGrath JM, Funk A, Galewski P, Ou S, Townsend B, Davenport K, et al. A contiguous *de novo* genome assembly of sugar beet EL10 (*Beta vulgaris* L.). *DNA Res*. 2023;30:dsac033.
38. Sielemann K, Pucker B, Orsini E, Elashry A, Schulte L, Viehoveer P, et al. Genomic characterization of a nematode tolerance locus in sugar beet. *BMC Genomics*. 2023;24:748.
39. Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2013–2015. <http://www.repeatmasker.org>.

Publisher's Note

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