frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 26 June 2023
DOI 10.3389/fpls.2023.1193122

Check for updates

OPEN ACCESS

EDITED BY
Ute Vothknecht,
University of Bonn, Germany

REVIEWED BY
Tiago Santana Balbuena,
São Paulo State University, Brazil
Ilka Wittig,
Goethe University Frankfurt, Germany

*CORRESPONDENCE
Hans-Peter Braun
braun@genetik.uni-hannover.de

RECEIVED 24 March 2023
ACCEPTED 02 June 2023
PUBLISHED 26 June 2023

CITATION
Schröder L, Rupp O, Senkler M, Rugen N, Hohnjec N, Goesmann A, Küster H and Braun H-P (2023) The *Viscum album* Gene Space database.
Front. Plant Sci. 14:1193122.
doi: 10.3389/fpls.2023.1193122

COPYRIGHT
© 2023 Schröder, Rupp, Senkler, Rugen, Hohnjec, Goesmann, Küster and Braun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The *Viscum album* Gene Space database

Lucie Schröder [1], Oliver Rupp [2], Michael Senkler [1], Nils Rugen [1], Natalija Hohnjec [3], Alexander Goesmann [2], Helge Küster [3] and Hans-Peter Braun [1*]

[1]Plant Proteomics, Institute of Plant Genetics, Leibniz Universität Hannover, Hannover, Germany, [2]Bioinformatics and Systems Biology, Justus-Liebig-Universität Gießen, Gießen, Germany, [3]Plant Genomics, Institute of Plant Genetics, Leibniz Universität Hannover, Hannover, Germany

The hemiparasitic flowering plant *Viscum album* (European mistletoe) is known for its very special life cycle, extraordinary biochemical properties, and extremely large genome. The size of its genome is estimated to be 30 times larger than the human genome and 600 times larger than the genome of the model plant *Arabidopsis thaliana*. To achieve insights into the Gene Space of the genome, which is defined as the space including and surrounding protein-coding regions, a transcriptome project based on PacBio sequencing has recently been conducted. A database resulting from this project contains sequences of 39,092 different open reading frames encoding 32,064 distinct proteins. Based on 'Benchmarking Universal Single-Copy Orthologs' (BUSCO) analysis, the completeness of the database was estimated to be in the range of 78%. To further develop this database, we performed a transcriptome project of *V. album* organs harvested in summer and winter based on Illumina sequencing. Data from both sequencing strategies were combined. The new *V. album* Gene Space database II (VaGs II) contains 90,039 sequences and has a completeness of 93% as revealed by BUSCO analysis. Sequences from other organisms, particularly fungi, which are known to colonize mistletoe leaves, have been removed. To evaluate the quality of the new database, proteome data of a mitochondrial fraction of *V. album* were re-analyzed. Compared to the original evaluation published five years ago, nearly 1000 additional proteins could be identified in the mitochondrial fraction, providing new insights into the Oxidative Phosphorylation System of *V. album*. The VaGs II database is available at https://viscumalbum.pflanzenproteomik.de/. Furthermore, all *V. album* sequences have been uploaded at the European Nucleotide Archive (ENA).

KEYWORDS

database development, PacBio sequencing, Illumina sequencing, Complexome profiling, mitochondria, oxidative phosphorylation (OXPHOS), complex I, supercomplex

Introduction

European mistletoe (*V. album*) is an obligate hemiparasitic flowering plant of the order Santalales. It grows on numerous trees in Europe. The host trees provide *V. album* with water, nutrients and, to a certain extent, with organic compounds. At the same time, *V. album* performs photosynthesis and produces organic compounds itself. In contrast to most angiosperms in Central Europe, it does not discard its leaves in winter and performs photosynthesis all year. The vitality of host trees may be impaired by mistletoe settlement.

V. album is known for a very special lifestyle (see Glatzel and Geils, 2009 for review): *V. album* does not germinate in soil but on branches of trees where it becomes connected with the xylem of its host. To ensure spreading, the fruits of *V. album* ripe in winter, when other food resources are scarce for birds. The fruits are very sticky to ensure a stable attachment on branches. The seeds lack a seed coat and consist of an embryo, which can germinate from the fruit without a dormancy phase. Haustoria, which are formed first during germination, are guided *via* negative phototropism to the surface of the branch of the host tree. After connection with the xylem of the vascular system, the haustoria take up water, minerals and organic compounds from the host plants. Afterwards, one pair of shoot segments per year per shoot apical meristem are formed by the *V. album* plant. The typical ball-like shape is achieved after several years by annual realignment of the new shoots, which grow in all directions. In contrast to most other plant species the leaves of *V. album* do not close their stomata during water shortage, which can increase water stress of the host plant. In August/September leaves of the previous year are discarded, without recycling chlorophyll, while the new leaves formed in spring stay attached, to perform photosynthesis in the winter.

Biochemically, *V. album* stands out with its rich content of phenolic acids, phenylpropanoids, flavonoids, triterpenes and phytosterols (Urech and Baumgartner, 2015; Jäger et al., 2021). Furthermore, *V. album* produces specific proteins, the viscotoxins and mistletoe lectins, which act as a biotic defense system. The stickiness of the fruits is provided by special kinds of hemicellulose compounds (Azuma Ji et al., 2000). The development of *V. album* is controlled by an extraordinary distribution of phytohormones. *V. album* extracts have immune stimulating and cytotoxic effects, which are used in medicine (Nazaruk and Orlikowski, 2016).

At the molecular level, *V. album* is less well characterized. The mitochondrial and chloroplast genomes have been sequenced (Petersen et al., 2015a; Petersen et al., 2015b; Skippington et al., 2015; Skippington et al., 2017) and found to lack some of the genes normally present in these organelles, especially those encoding subunits of the mitochondrial NADH dehydrogenase complex and the homologous chloroplastidic NDH complex. The complete absence of these complexes was shown by proteomic studies (Maclean et al., 2018; Senkler et al., 2018; Schröder et al., 2020; Schröder et al., 2022a).

The genome of *V. album* consists of 2n=20 chromosomes and is considered to be one of the larges genomes of flowering plants (Zonneveld, 2010; Novák et al., 2020). It consists of almost 100 billion base pairs. More than 50% of the genome sequence of *V. album* consists of highly repetitive DNA (Novák et al., 2020). The genome sequence of *V. album* has not been determined to date but the partial sequence of its Gene Space has recently been presented (Schröder et al., 2022b). The GC content of the gene sequences lies at about 50%, which is exceptionally high for angiosperms.

An initial approach to analyze the *V. album* Gene Space was based on Single Molecule Real-Time (SMRT) sequencing (PacBio sequencing) (Schröder et al., 2022b). A database resulting from this project contains 39,092 gene sequences, from which 32,064 protein sequences were derived. The results enabled the development of a first *V. album* Gene Space database. The completeness of this database was estimated to be in the range of 78%. To further develop this database, a *V. album* transcriptome project has been carried out using the Illumina sequencing approach. Analyses were performed for *V. album* samples harvested in winter and summer, respectively. By combining novel and existing sequencing data, we here present a new *V. album* database including 90,039 sequence entries. The quality of the new database is demonstrated by re-evaluation of published *V. album* proteome datasets.

Materials and methods

Isolation of mRNA fractions from *V. album*

Mistletoes (European mistletoe; *Viscum album*), grown on an apple tree (*Malus* sp.) on the campus of Leibniz University Hannover were used as starting material. Various organs (leaves, stems, flower buds) of male and female plants were harvested in summer and in winter, shock-frozen using liquid nitrogen, and stored at –80°C until use. mRNA isolation and quality evaluation were carried out as described previously (Schröder et al., 2022b).

Sequence analysis

mRNA fractions were reverse transcribed into cDNA. The summer and winter fractions were analyzed separately using Illumina PE150 (paired-end read) sequencing. cDNA libraries containing 250~300 bp inserts were constructed and sequenced. Quantitative data as well as quality evaluation data of the libraries are given in Table 1.

Transcriptome assembly

The raw Illumina RNA-seq reads were first trimmed for low-quality and adaptor regions using Trimmomatic (Bolger et al., 2014) (version 0.36, ILLUMINACLIP : ADAPTER.fa:2:20:7 SLIDINGWINDOW:4:15 MINLEN:50). The trimmed reads were assembled using Trinity (Haas et al., 2013) (version 2.11.0) including *in silico* normalization with a target coverage of 50 x. The PacBio IsoSeq reads were included in the assembly using the "–long_reads" option of Trinity. The assembled transcripts were screened for coding sequences using TransDecoder (Haas and Papanicolaou, n.d.). The potential protein sequences were merged

TABLE 1 Results and quality evaluation of the Illumina sequencing approach of the winter and summer samples of *V. album*.

Sample[a]	Raw reads[b]	Raw data(G)[c]	Error (%)[d]	Q20 (%)[e]	Q30 (%)[f]	GC (%)[g]
winter	120,720,364	18,108,054,600	0.03	97.96	93.99	49.47
summer	121,840,742	18,276,111,300	0.02	98.11	94.34	49.37

[a]sample names.
[b]original sequencing reads counts.
[c]raw reads number multiplied by read lengths, saved in G unit.
[d]average sequencing error rate, calculated by Qphred=-10log10(e).
[e]percentages of bases whose correct base recognition rates are greater than 99% in total bases.
[f]percentages of bases whose correct base recognition rates are greater than 99.9% in total bases.
[g]percentages of G and C in total bases.

with the previously created protein sequences (Schröder et al., 2022b) and redundant sequences were filtered using CD-HIT (Li and Godzik, 2006) (version 4.7, -c 0.95 -aS 0.99). The functional annotation was computed using the best blast hit method on the UniProt/Swiss-Prot (Boutet et al., 2007) and UniProt/TrEMBL (The UniProt Consortium et al., 2023) databases. The TPM values were computed with Salmon (Patro et al., 2017) (version 1.9.0). The completeness of the filtered transcripts was estimated with BUSCO (Seppey et al., 2019) (version 5.0.0) using the viridiplantae database.

The *V. album* database

The final *V. album* Gene Space II database was created by combining PacBio and Illumina sequencing. It includes 90,039 distinct proteins. The database is accessible at https://viscumalbum.pflanzenproteomik.de/. Furthermore, all sequences were submitted to NCBI.

Re-evaluation of the *V. album* mitochondria complexome

The complexome profiling approach is introduced in Schröder et al., 2022c. Primary complexome profiling data were taken from Senkler et al., 2018. Re-evaluation of the mass-spectrometry (MS) data and annotation of proteins were carried out with MaxQuant (version 2.1.4.0) using the novel VaGs II database. The settings were the same as in Senkler et al., 2018. For heatmap generation, the abundance profiles (based on iBAQ values; Schwanhäusser et al., 2011; calculated by MaxQuant) of all identified proteins were used. These profiles were aligned according to similarity by hierarchical clustering using the NOVA software (version: 0.5.8; Giese et al., 2015). The Complexome Map of the re-evaluated *V. album* mitochondria fractions is available at the ComplexomeMap portal at https://complexomemap.de/75.

Results and discussion

For Illumina analysis, transcripts from leaves, stems and buds of female and male *V. album* plants were isolated, combined and reverse-transcribed into cDNA. Organs were harvested in winter and summer and corresponding cDNA fractions were analyzed separately. Illumina sequence analysis revealed >120,000,000 reads for the summer and the winter fraction, respectively (Table 1).

Transcriptome assembly

The datasets were processed as described in the Materials and Methods section and assembled using the Trinity software package version 2.11.0 (Haas et al., 2013). The PacBio IsoSeq reads (Schröder et al., 2022b) were included to the assembly using the "–long_reads" option of Trinity. The Trinity assembly produced 650,594 transcript contigs. TransDecoder predicted potential coding sequences on 144,517 contigs. The clustering with CD-HIT produced 104,405 unique sequences, which were used for further analyses.

Functional annotation of transcripts and functional evaluation of the *V. album* transcript samples

Functional annotation of the 104,405 transcripts was computed using the best blast hit method on the UniProt/Swiss-Prot (Boutet et al., 2007) and UniProt/TrEMBL (The UniProt Consortium et al., 2023) databases. As a result, 51% of the sequences could be assigned a function. As expected, in most cases the greatest similarity is with sequences from the phylum *Viridiplantae* (green algae and the land plants). However, there are also numerous transcripts which show highest similarity to transcripts from other phyla, especially fungi. No functions could be assigned to many other sequences, which can likely be explained by the fact that no genome sequences of plants more closely related to *V. album* are yet available.

V. album organs used for transcript isolation were from plants living in the field. *V. album* leaves are known to be colonized by several endophytic fungi (Peršoh et al., 2010; Kotan et al., 2013, Ariantari et al. 2019, reviewed in Krasylenko et al., 2020). To investigate the presence of fungal sequences in our database, we evaluated the functional annotation of Illumina transcript sequences according to Phyla for the winter and summer samples (Figure 1). 37.614 of the *V. album* transcripts assigned to *Viridiplantae*. 91% of them (34.159 transcripts) likewise were present in both, the winter and the summer sample. Furthermore, 11.830 transcripts of our

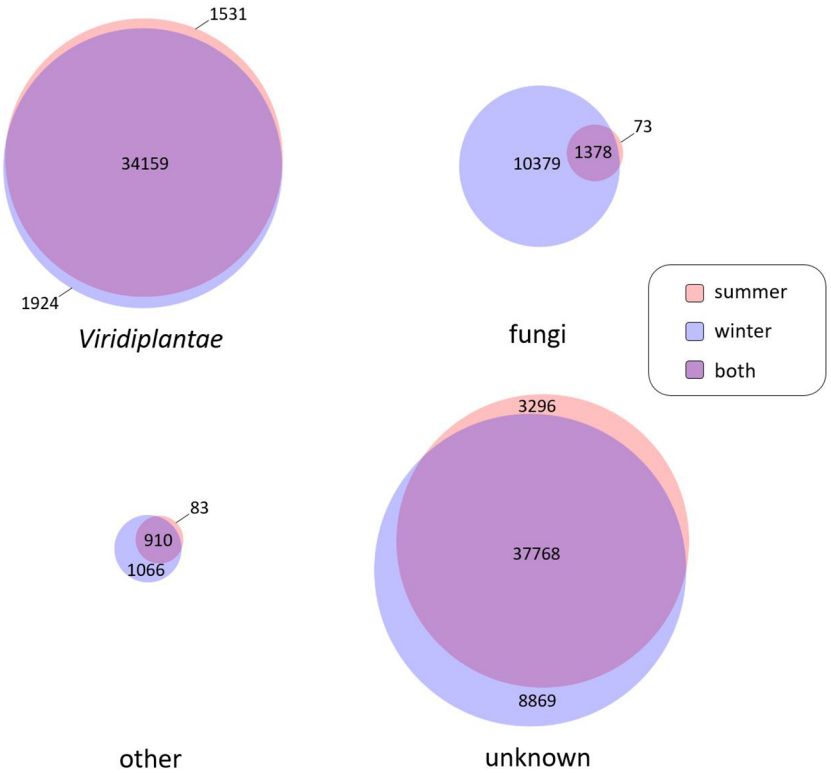

FIGURE 1
Number of transcripts according to assignment to phyla. Assignment of transcripts obtained by Illumina Sequencing was based on the best blast hit method on the UniProt/Swiss-Prot and UniProt/TrEMBL databases. Number of transcripts of summer and winter samples were displayed separately. Phyla considered are *Viridiplantae* (green algae and the land plants (embryophytes)) and fungi (as yeasts, molds and mushrooms). Transcripts resembling those of other phyla (e.g.: metazoan, bacteria etc.) were grouped under "other". Transcripts not significantly resembling known sequences were labeled "unknown". The Venn diagram was created by BioVenn (Hulsen et al., 2008).

V. album samples were assigned to the fungi Phylum (Figure 1). Almost all fungal sequences are from the winter sample, which is probably due to the increased age of the mistletoe leaves (new leaves appear in the spring and are kept in the winter). Only 11% of the fungal sequences occur likewise in the summer and winter samples. Few of the transcripts of our *V. album* samples were assigned to other Phyla than *Viridiplantae* or fungi (Figure 1). Transcripts that could not be functionally annotated were labeled "unknown" (Figure 1). The number of transcripts of this category is similar in the winter and summer samples; also, they show quite a high intersection in terms of their occurrence in the two seasons (76%), similar to the transcripts assigned to *Viridiplantae*. We conclude that these sequences should be predominantly *V. album* transcripts. The relatively high proportion of transcripts of unknown function likely is due to the lack of comparative sequences from related plant species, but also due to the fact that *V. album* has a special way of life, which requires numerous proteins that do not occur in other phyla.

To further analyze the origin of sequences in our *V. album* samples, we quantified all transcripts identified by our Illumina sequencing approach according to their assignment to Phyla (Figure 2). Transcripts per million (TPM) values of transcripts in the summer and winter samples were examined separately. As expected, TMP values are highest for transcripts assigned to *Viridiplantae* and clearly lower in those assigned to the fungi Phylum. This is particularly visible in the sample harvested in summer. TMP values of unassigned transcripts more resemble the values obtained for transcripts assigned to *Viridiplantae*, which is especially evident in the summer sample. This reconfirms that the transcripts in this category are derived from *V. album* rather than from fungi colonizing *V. album*. The absolute summed up TPM values (summer and winter sample combined) for the four categories are 1,096,579 for *Viridiplantae* (73%), 385,858 for "unknown" (25%), 11,497 for fungi (0.8%) and 17,789 for "others" (1.1%). Thus, about 1% of the transcripts of our *V. album* fractions can be classified as fungal. This should realistically reflect the conditions in the "ecological niche" *V. album*.

The *Viscum album* Gene Space database II

The previous *V. album* Gene Space database, which was created based on Single Molecule Real-Time (SMRT) sequencing (PacBio sequencing), includes 39,092 entries encoding 32,064 distinct proteins (Schröder et al., 2022b). We now designate this database VaGs I. BUSCO analyses had revealed a completeness of about 78%. Through implementation of the new Illumina sequencing data and

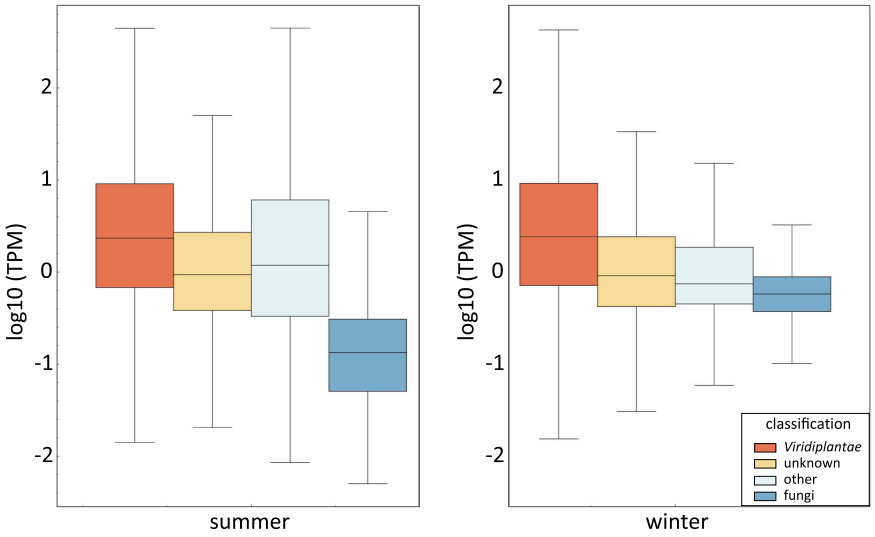

FIGURE 2
Transcript levels within the *V. album* summer and winter samples according to their assignment of phyla. Transcript levels are given in log10 Transcripts Per Million (TPM). The line within the box shows the median for each dataset. The upper end of the box is the "upper quartile", which is the median of the upper half of the dataset. The lower quartile (lower end of the box) is the median of the lower half of the dataset. The end of the vertical lines indicate the lowest and highest values of the dataset. The diagram was created by instant clue (Nolte et al., 2018).

improved transcript annotation, we hereby present a novel *V. album* Gene Space database, VaGs II. We decided to remove transcripts that are unambiguously assigned to the Phyla Fungi, Animals, Bacteria and Viruses based on sequence comparisons. VaGs II includes all transcripts assigned to the categories "*Viridiplantae*" and "unknown". The totel number of distinct transcripts is 90,039. The BUSCO score of VaGs II is 93% (91% complete, 2% fragmented; Figure 3).

A database has been developed that allows all sequences to be viewed (Figure 4). Sequences can be downloaded, proteins can be searched by ID, name, and sequence, and instructions for BLAST searches are given. Furthermore, all *V. album* transcript have been uploaded at the European Nucleotide Archive (ENA) (https://www.ebi.ac.uk/ena/browser/home).

Re-evaluation of complexome profiling data for *V. album* mitochondria using VaGs II

To test the quality of the new VaGs II databank, a published proteomic dataset on *V. album* mitochondria (Senkler et al., 2018) was re-evaluated. This is a complexome dataset which was created as follows: First, mitochondria were isolated from *V. album* and a mild detergent was used to dissolve the mitochondrial membranes. Protein complexes in the resulting solution were separated by Blue-native polyacrylamide gel electrophoresis. A lane of the native gel was then cut from top to bottom into 54 small gel slices, which were individually subjected to label-free quantitative shotgun mass spectrometry for protein identifications. Finally, abundance

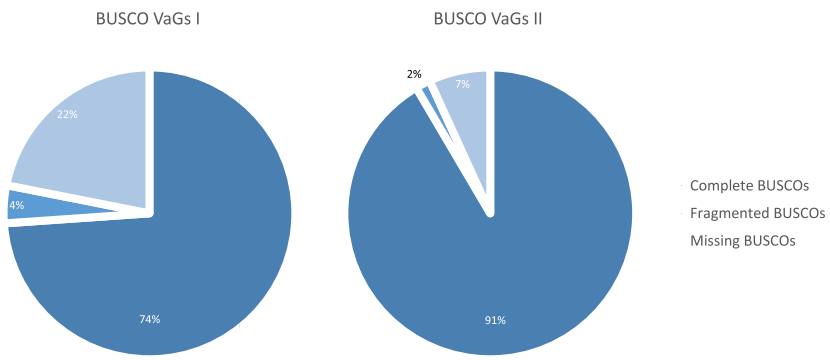

FIGURE 3
Comparison of the completeness of the VaGs I database (Schröder et al., 2022b) and the new VaGs II database as revealed by 'Benchmarking Universal Single Copy Orthologs' (BUSCO) analysis (Seppey et al., 2019).

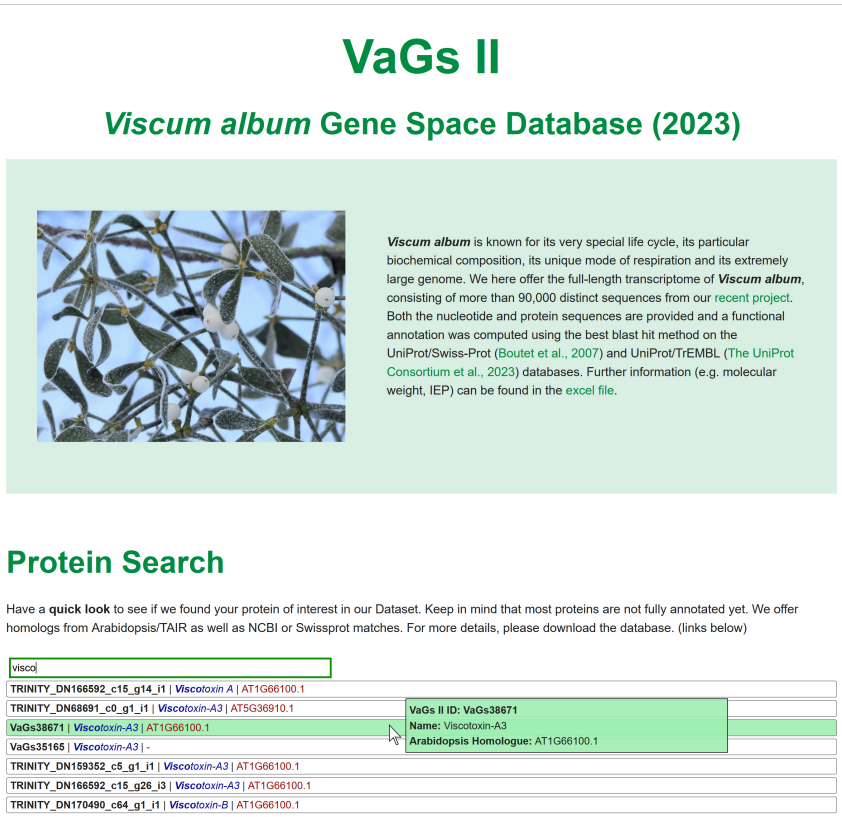

FIGURE 4
The *Viscum album* Gene space database at https://viscumalbum.pflanzenproteomik.de/.

profiles along the blue-native gel lane were calculated for all identified proteins. The results were visualized as a heatmap and the profiles of individual proteins were aligned according to similarity. On the resulting heatmap, proteins belonging to the same protein complex form characteristic clusters. It thus is possible to systematically investigate the protein complexes of a biochemical fraction, the 'complexome'. The original data of this experiment still had to be analyzed using a protein database for the model plant *Arabidopsis thaliana* (Senkler et al., 2018), since a *V. album* database was not available. This allowed the identification of 477 proteins in total. Re-evaluation of the same dataset using the VaGs II database now allowed the identification of 1392 proteins (Figure 5). The newly evaluated mitochondrial complexome of *V. album* is presented as a supplement of this publication

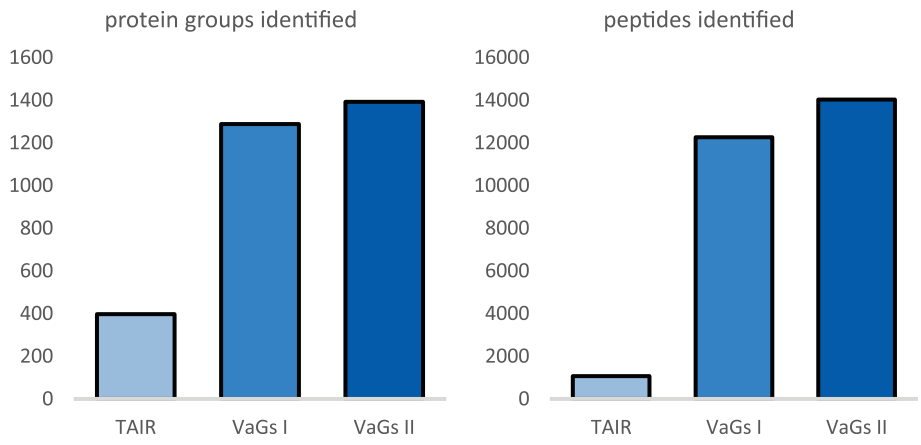

FIGURE 5
Number of identified proteins or peptides in a *V. album* mitochondrial complexome dataset upon data evaluation with TAIR, VaGs I and VaGs II. TAIR10: Arabidopsis protein database provided by The *Arabidopsis thaliana* Information resource (TAIR), Version 10 (https://www.arabidopsis.org/). The mitochondrial complexome dataset is derived from Senkler et al., 2018. The dataset was evaluated by TAIR10 (as in Senkler et al., 2018), VaGs I (Schröder et al., 2022b) and VaGs II (this study).

10.3389/fpls.2023.1193122

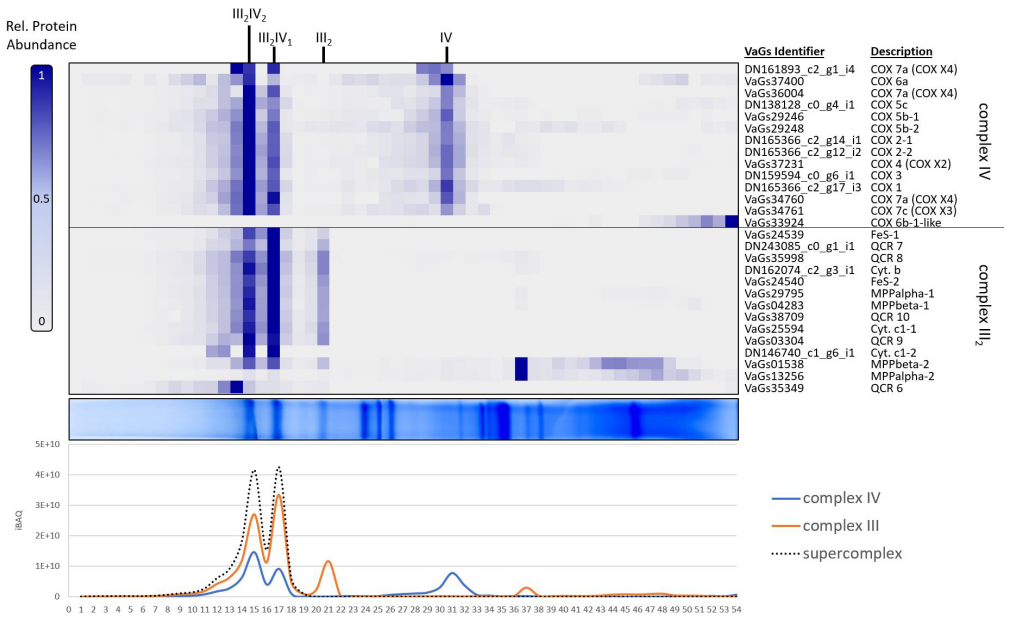

FIGURE 6
Identified subunits of complexes III and IV of the respiratory chain from *V. album* in a mitochondrial complexome dataset (Senkler et al., 2018) upon analysis using the VaGs II database. Mitochondrial proteins were separated by Blue native (BN) PAGE and stained by Coomassie-blue (horizontal gel lane in the center of the Figure). The gel lane was dissected into 54 slices, which all were subjected to label-free quantitative shotgun mass spectrometry. Abundance profiles of individual proteins along the gel lane are visualized as a heatmap (figure part above the gel lane; the columns correspond to the 54 gel slices; the rows to the abundance profiles of individual proteins; normalized (max) abundance profiles are given; see scale to the left of the heatmap) and aligned according to similarity (using the NOVA software package version 0.5.8). Only profiles of the subunits of respiratory complexes III and IV are shown (for complete dataset see Supplementary Data S1 and ComplexomeMap at https://complexomemap.de/75). Protein identities are given to the right of the heatmap. Summed-up abundance profiles (given as 'intensity based absolute quantification' (iBAQ) values) of the subunits of the two protein complexes as well as assemblies of the complexes (supercomplexes III_2IV_1 and III_2IV_2) are given in the diagram (figure part below the gel lane; y-axis: iBAQ score; x-axis: gel slices 1-54). Results of the original evaluation (Senkler et al., 2018) are shown in Supplementary Figure S1).

(Supplementary Table 1) and is also accessible at the ComplexomeMap portal at https://complexomemap.de/75. It offers novel insights into the molecular biology of *V. album* mitochondria. Also, the number of peptides identified in the complexome fractions (gel slices) significantly increased based on VaGs II evaluation, resulting in a better coverage of proteins by peptides.

To further analyze the mitochondrial complexome from *V. album*, as revealed by VaGs II evaluation, we took a closer look at the protein clusters corresponding to respiratory chain complexes III and IV (Figure 6). Complex III is an ubiquinol:cytochrome-c oxidoreductase and complex IV a cytochrome-c:O_2 oxidoreductase. These two protein complexes catalyze the last two steps of the mitochondrial respiratory chain and were found to form two exceptionally stable supercomplexes in *V. album* mitochondria (Senkler et al., 2018). According to current knowledge (Maldonado et al., 2021), complexes III and IV in plants each consist of 10 different subunits (the 10 subunits of complex III, however, all are present in duplicate, as complex III occurs as a functional dimer). In *V. album*, this dimer (III_2) associates with one or two copies of complex IV; the corresponding supercomplexes are designated III_2IV_1 and III_2IV_2. Of the total 20 different proteins present in these supercomplexes, all 20 proteins were found in our complexome dataset upon evaluation using VaGs II (Figure 6; compared to only 8 proteins based on the original evaluation, Senkler et al., 2018; see Supplementary Figure S1). 18 of the identified subunits form part of the two supercomplexes, whereas two (QCR6 of complex III_2 and Cox6b from complex IV) are partly detached and migrate in the low-molecular-mass region of the blue native gel, probably because they became detached during membrane solubilization. For some of the proteins, different isoforms are present in *V. album* (e.g. the Cox5b and the α- and β-MPP subunits, Supplementary Figures S2, S3). The summed-up abundance profiles of the subunits of complex III and of complex IV reveal that the amounts of the two supercomplexes are similar; furthermore, the stoichiometry of the two complexes within the supercomplexes is correctly revealed (Figure 6, bottom).

Concluding remarks

We present here the *Viscum album* Gene Space database II, VaGs II. Based on our quantitative and qualitative evaluations, we assume that it covers well above 90% of the protein-coding genes of *V. album*. The database has been cleaned with respect to sequences from other organisms, particularly fungi that are known to colonize *V. album* leaves. The database contains 90,039 transcript sequences. In particular, the functional annotation of the sequences was

improved. Moreover, winter and summer samples of *V. album* were studied separately based on Illumina sequencing. Further analysis of these data should provide insights into adaptations of *V. album* to different seasons. VaGs II should provide an important data background useful for further studying the molecular biology of this extraordinary plant.

Data availability statement

The new VaGs II database is available at https://viscumalbum.pflanzenproteomik.de/. In addition, fasta files of thenucleotide or protein sequences can be downloaded. All Illumina and PacBio sequencing datasets and their assembly dataset have been uploaded to the European Nucleotide Archive (ENA) and can be found under the study identifier PRJEB60149. Within this study, the PacBio dataset can be found under the identifiers ERR10970196 and ERR10970197, the Illumina summer dataset under ERR10968077 and the Illumina winter dataset under ERR10968073. The original mass spectrometry proteomics data (Senkler et al., 2018) have been deposited to the ProteomeXchange Consortium via the PRoteomics IDEntifications database (PRIDE, Perez-Riverol et al., 2022) partner repository with the dataset identifier PXD008974. The re-evaluated dataset based on an evaluation using the new VaGs II database has been deposited additionally to the ProteomeXchange Consortium via PRIDE and is accessible under PXD041061.

Author contributions

HPB, HK, and LS initiated and supervised the project. NH isolated mRNA from *V. album* organs. OR, AG, and LS carried out database development and data annotation. NR and LS re-evaluated complexome-profiling data using the new database. MS developed the web portal for the *V. album* Gene Space database. LS and HPB performed data evaluation and interpretation. LS and HPB wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research has been supported by the Deutsche Forschungsgemeinschaft, grant BR 1829/16-1, to HPB. The publication of this article was funded by the Open Access Fund of Leibniz Universität Hannover.

Acknowledgments

Illumina sequencing was carried out by Novogene, Cambridge, UK.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1193122/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Identified subunits of complexes III and IV of the respiratory chain from V. album in a mitochondrial complexome dataset (Senkler et al., 2018) upon analysis using the Arabidopsis TAIR10 database (https://www.arabidopsis.org/). The same dataset was re-analyzed using the VaGs database (Figure 6).

SUPPLEMENTARY FIGURE 2
Sequence alignment of *V. album* COX5b-1 and COX5b-2. The COX 5b subunit of the cytochrome c oxidase complex (Complex IV) is present in two isoforms, COX5b-1 and COX5b-2. The proteins are encoded by transcripts VaGs29248 and VaGs29246. Conserved amino acids positions are highlighted in blue. Both proteins form part of the III_2IV and III_2IV_2 supercomplexes in the mitochondrial complexome dataset (Figure 4 and Supplementary Data S1).

SUPPLEMENTARY FIGURE 3
Identified pairs of isoforms of the two subunits of the mitochondrial processing peptidase (MPP) from *V. album* in a mitochondrial complexome dataset (Senkler et al., 2018) upon re-evaluation using the VaGs II database. Both, the α- and the β-MPP subunits form part of the cytochrome c reductase complex (complex III) in plants (Braun, 2020). The β-subunit is encoded by transcripts VaGs01538 and VaGs04283; the α-subunit by transcripts VaGs13256 and VaGs29795. The heatmap indicates the abundance profiles of the four proteins along a blue-native gel lane used for the complexome profiling analysis. The positions of complexes III and IV and their supercomplexes are indicated above the heatmap, the molecular masses of standard proteins below the heatmap. One isoform of the α- and the β-subunit of MPP forms part of the III_2IV and III_2IV_2 supercomplexes, respectively. The other two isoforms seem to (partially) form a separate protein complex in the 110-160 kDa range, which might represent a soluble form of the heterodimeric MPP enzyme (detached from complex III). A soluble MPP enzyme has been described for mitochondrial matrix of animals and yeast (Braun and Schmitz, 1995), but possibly also occurs in *V. album*.

SUPPLEMENTARY DATA SHEET 1
Heatmap of abundance profiles of the proteins included in the mitochondrial complexome dataset of *V. album* re-evaluated using the VaGs II database.

References

Ariantari, N. P., Daletos, G., Mándi, A., Kurtán, T., Müller, W. E. G., Lin, W., et al. (2019). Expanding the chemical diversity of an endophytic fungus *Bulgaria inquinans*, an ascomycete associated with mistletoe, through an OSMAC approach. *RSC Adv.* 9 (43), 25119–25132. doi: 10.1039/C9RA03678D

Azuma Ji, K. N. H., Heux, L., Vuong, R., and Chanzy, H. (2000). The cellulose system in viscin from mistletoe berries. *Cellulose* 7, 3–19. doi: 10.1023/A:1009223730317

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., and Bairoch, A. (2007). "UniProtKB/Swiss-prot," in *Plant bioinformatics, methods in molecular BiologyTM*. Ed. D. Edwards (Humana Press), 89–112.

Braun, H. P. (2020). The oxidative phosphorylation system of the mitochondria in plants. *Mitochondrion* 53, 66–75. doi: 10.1016/j.mito.2020.04.007

Braun, H. P., and Schmitz, U. K. (1995). The bifunctional cytochrome c reductase/processing peptidase complex from plant mitochondria. *J. Bioenerg Biomembr* 27 (4), 423–436. doi: 10.1007/BF02110005

Giese, H., Ackermann, J., Heide, H., Bleier, L., Dröse, S., Wittig, I., et al. (2015). NOVA: a software to analyze complexome profiling data. *Bioinformatics* 31, 440–441. doi: 10.1093/bioinformatics/btu623

Glatzel, G., and Geils, B. W. (2009). Mistletoe ecophysiology: host–parasite interactions. *Botany* 87, 10–15. doi: 10.1139/B08-096

Haas, B. J., and Papanicolaou, A. *TransDecoder (Find coding regions within transcripts)*. Available at: https://transdecoder.github.io/ (Accessed 3.30.17).

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084

Hulsen, T., de Vlieg, J., and Alkema, W. (2008). BioVenn – a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* 9, 488. doi: 10.1186/1471-2164-9-488

Jäger, T., Holandino, C., Melo, M. N. O., Peñaloza, E. M. C., Oliveira, A. P., Garrett, R., et al. (2021). Metabolomics by UHPLC-Q-TOF reveals host tree-dependent phytochemical variation in *Viscum album* l. *Plants (Basel)* 10 (8), 1726. doi: 10.3390/plants10081726

Kotan, R., Okutucu, A., Ala Görmez, A., Karagoz, K., Dadasoglu, F., Karaman, İ, et al. (2013). Parasitic bacteria and fungi on common mistletoe (*Viscum album* l.) and their potential application in biocontrol. *J. Phytopathol.* 161, 165–171. doi: 10.1111/jph.12048

Krasylenko, Y., Sosnovsky, Y., Atamas, N., Popov, G., Leonenko, V., Janošíková, K., et al. (2020). The European mistletoe (Viscum album l.): distribution, host range, biotic interactions, and management worldwide with special emphasis on Ukraine. *Botany*. 98 (9), 499–516. doi: 10.1139/cjb-2020-0037

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinforma. Oxf. Engl.* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158

Maclean, A. E., Hertle, A. P., Ligas, J., Bock, R., Balk, J., and Meyer, E. H. (2018). Absence of complex I is associated with diminished respiratory chain function in European mistletoe. *Curr. Biol.* 28 (10), 1614–1619.e3. doi: 10.1016/j.cub.2018.03.036

Maldonado, M., Guo, F., and Letts, J. A. (2021). Atomic structures of respiratory complex III_2, complex IV, and supercomplex III_2-IV from vascular plants. *eLife* 10, e62047. doi: 10.7554/eLife.62047

Nazaruk, J., and Orlikowski, P. (2016). Phytochemical profile and therapeutic potential of viscum album l. *Nat. Prod Res.* 30 (4), 373–385. doi: 10.1080/14786419.2015.1022776

Nolte, H., MacVicar, T. D., Tellkamp, F., and Krüger, M. (2018). Instant clue: a software suite for interactive data visualization and analysis. *Sci. Rep.* 8, 12648. doi: 10.1038/s41598-018-31154-6

Novák, P., Guignard, M. S., Neumann, P., Kelly, L. J., Mlinarec, J., Koblížková, A., et al. (2020). Repeat-sequence turnover shifts fundamentally in species with large genomes. *Nat. Plants.* 6 (11), 1325–1329. doi: 10.1038/s41477-020-00785-x

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. doi: 10.1038/nmeth.4197

Perez-Riverol, Y., Bai, J., Bandla, C., Hewapathirana, S., García-Seisdedos, D., Kamatchinathan, S., et al. (2022). The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* 50 (D1), D543–D552. doi: 10.1093/nar/gkab1038

Peršoh, D., Melcher, M., Flessa, F., and Rambold, G. (2010). First fungal community analyses of endophytic ascomycetes associated with viscum album ssp. *austriacum its Host Pinus sylvestris. Fungal Biol.* 114 (7), 585–596. doi: 10.1016/j.funbio.2010.04.009

Petersen, G., Cuenca, A., Møller, I. M., and Seberg, O. (2015a). Massive gene loss in mistletoe (Viscum, viscaceae) mitochondria. *Sci. Rep.* 5, 17588. doi: 10.1038/srep17588

Petersen, G., Cuenca, A., and Seberg, O. (2015b). Plastome evolution in hemiparasitic mistletoes. *Genome Biol. Evol.* 7 (9), 2520–2532. doi: 10.1093/gbe/evv165

Schröder, L., Eubel, H., and Braun, H. P. (2022c). Complexome profiling of plant mitochondrial fractions. *Methods Mol. Biol.;* 2363, 101–110. doi: 10.1007/978-1-0716-1653-6_9

Schröder, L., Hegermann, J., Pille, P., and Braun, H. P. (2022a). The photosynthesis apparatus of European mistletoe (*Viscum album*). *Plant Physiol.* 190 (3), 1896–1914. doi: 10.1093/plphys/kiac377

Schröder, L., Hohnjec, N., Senkler, M., Senkler, J., Küster, H., and Braun, H. P. (2022b). The gene space of European mistletoe (*Viscum album*). *Plant J.;* 109 (1), 278–294. doi: 10.1111/tpj.15558

Schröder, L., Senkler, J., and Braun, H. P. (2020). "Special features of cellular respiration in viscum album," in *Die mistel in der tumortherapie 5*. Eds. R. Scheer, et al (Essen: KVC Verlag), pp 3–pp13.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., et al. (2011). Global quantification of mammalian gene expression control. *Nature.* 473 (7347), 337–342. doi: 10.1038/nature11848

Senkler, J., Rugen, N., Eubel, H., Hegermann, J., and Braun, H. P. (2018). Absence of complex I implicates rearrangement of the respiratory chain in European mistletoe. *Curr. Biol.* 28 (10), 1606–1613.e4. doi: 10.1016/j.cub.2018.03.050

Seppey, M., Manni, M., and Zdobnov, E. M. (2019). "BUSCO: assessing genome assembly and annotation completeness," in *Gene prediction. methods in molecular biology*, vol. 1962 . Ed. M. Kollmar (New York, NY: Humana), 2019. doi: 10.1007/978-1-4939-9173-0_14

Skippington, E., Barkman, T. J., Rice, D. W., and Palmer, J. D. (2015). Miniaturized mitogenome of the parasitic plant viscum scurruloideum is extremely divergent and dynamic and has lost all nad genes. *Proc. Natl. Acad. Sci. U S A.* 112 (27), E3515–E3524. doi: 10.1073/pnas.1504491112

Skippington, E., Barkman, T. J., Rice, D. W., and Palmer, J. D. (2017). Comparative mitogenomics indicates respiratory competence in parasitic viscum despite loss of complex I and extreme sequence divergence, and reveals horizontal gene transfer and remarkable variation in genome size. *BMC Plant Biol.* 17 (1), 49. doi: 10.1186/s12870-017-0992-8

The UniProt Consortium, Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Ahmad, S., et al. (2023). UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res.* 51, D523–D531. doi: 10.1093/nar/gkac1052

Urech, K., and Baumgartner, S. (2015). "Chemical constituents of viscum album l.: implications for the pharmaceutical preparation of mistletoe," in *Mistletoe: from mythology to evidence-based medicine*, vol. 4 . Eds. K. S. Zänker and S. V. Kaveri (Transl. Res. Biomed. Basel, Karger), pp 11–pp 23. doi: 10.1159/000375422

Zonneveld, B. J. M. (2010). New record holders for maximum genome size in eudicots and monocots. *J. Bot.* 2010, 527357, 4. doi: 10.1155/2010/527357