data

MDPI

Data Descriptor

NGS Reads Dataset of Sunflower Interspecific Hybrids

Maksim S. Makarenko [1,*] and Vera A. Gavrilova [2]

[1] The Laboratory of Plant Genomics, The Institute for Information Transmission Problems, 127051 Moscow, Russia
[2] The N.I. Vavilov All Russian Institute of Plant Genetic Resources, 190031 Saint Petersburg, Russia
* Correspondence: mcmakarenko@yandex.ru

Abstract: The sunflower (*Helianthus annuus*), which belongs to the family of Asteraceae, is a crop grown worldwide for consumption by humans and livestock. Interspecific hybridization is widespread for sunflowers both in wild populations and commercial breeding. The current dataset comprises 250 bp and 76 paired-end NGS reads for six interspecific sunflower hybrids (F1). The dataset aimed to expand *Helianthus* species genomic information and benefit genetic research, and is useful in alloploids' features investigations and nuclear–organelle interactions studies. Mitochondrial genomes of perennial sunflower hybrids *H. annuus* × *H. strumosus* and *H. annuus* × *H. occidentalis* were assembled and compared with parental forms.

Dataset: The National Center for Biotechnology BioProject: PRJNA929972 https://www.ncbi.nlm.nih.gov/bioproject/PRJNA929972.

Keywords: sunflower; *Helianthus*; interspecific hybrids; NGS data

Citation: Makarenko, M.S.; Gavrilova, V.A. NGS Reads Dataset of Sunflower Interspecific Hybrids. *Data* **2023**, *8*, 67. https://doi.org/10.3390/data8040067

Academic Editor: Juraj Gregáň

Received: 7 February 2023
Revised: 15 March 2023
Accepted: 24 March 2023
Published: 27 March 2023

1. Summary

The sunflower (*Helianthus annuus*), which belongs to the family of Asteraceae, is a crop that is grown worldwide for consumption by humans and livestock is also used in some industrial applications and as an ornamental in domestic gardens. The interspecific hybridization is widespread for sunflowers in nature, where it can lead to either the production of new subspecies or to the introgression of useful adaptive traits between species [1]. There is also great potential in agricultural systems to take advantage of this process for targeted crop improvement [2]. Wild *Helianthus* species are rich sources for genes determining resistance to different diseases, parasites, pests, drought, and other important traits [3]. Moreover, wild species may carry restoring fertility (*Rf*) genes, which are of potential interest for commercial hybrids (with high heterosis effect) production [4,5]. The present dataset comprises NGS reads of six interspecific sunflower hybrids. The dataset aimed to expand *Helianthus* species genomic information and benefit sunflower genetic studies.

2. Data Description

Here, we report NGS data for six interspecific sunflower hybrids (F1). Interspecific hybrids represent unique genetic material, especially those obtained between species with different ploidy. The current dataset includes more than 20.6 million 250 bp paired and 9.25 million 76 bp paired NGS reads. An example of reads quality analysis (FastQC data) is presented in Figure S1. The uncompressed data required more than 100 GB of disk space. The sequences have been deposited at National Center for Biotechnology Information (NCBI) SRA database (BioProject ID PRJNA929972). The sequence reads are stored in compressed files of FASTQ format with the following number of 250 bp paired-end reads for the samples: 3.01 mln—*H. annuus* (VIR100A) × *H. argophyllus* (1000),

8.45 mln—*H. annuus* (VIR114A) × *H. argophyllus* (1000), 1.86 mln—*H. annuus* (VIR100A) × *H. praecox* (560400), 2.81 mln—*H. annuus* (VIR117A) × *H. strumosus* (440679), 4.49 mln—*H. annuus* (VIR129A) × *H. occidentalis* (441062) and the following number of 76 bp paired-end reads: 3.97 mln—*H. annuus* (VIR129A) × *H. occidentalis* (441062) and 5.28 mln—*H. annuus* (HA89PET1)× *H. occidentalis* (441062).

Notably, the lowest (41%) GC content was mentioned in the hybrid combination *H. annuus* (VIR100A, VIR114A) × *H. argophyllus* (1000), while *H. argophyllus* is commonly used in crossing with *H. annuus* as a source of foreign genetic resources [6], and even such actions of crossing (*H. annuus* × *H. argophyllus*) were discovered in wild populations [7]. The highest GC content was 45%, in the case of hybridization with perennial species, which are quite rare viable progeny [8,9].

The data are insufficient for making nuclear genome assemblies. However, they may be used for investigations of plastid and mitochondrial genomes; the data are also appropriative for making variant (SNV) calling between subgenomes in the high copy regions of the nuclear genome, such as rDNA regions. Using current NGS reads data, we developed a complete mitochondrion assembly of two hybrids, *H. annuus* (VIR117A) × *H. strumosus* (440679) and *H. annuus* (VIR129A) × *H. occidentalis* (441062), which have predominantly perennial phenotypes.

In the case of the *H. annuus* (VIR117A) × *H. strumosus* (440679) hybrid, the size of the assembled mitochondrial genome was 305,217 bp; the *H. annuus* (VIR129A) × *H. occidentalis* (441062) mitogenome has 281,381 bp counts. Our previous studies investigated the mitochondrial genome structure of parental forms: maternal—*H. annuus* with PET1 type of cytoplasmic male sterility [10]—and paternal—*H. strumosus* and *H. occidentalis* [11,12]. Previous studies allowed us to compare the mitochondrial genome structure of the interspecific hybrid and its parental forms (Figure 1).

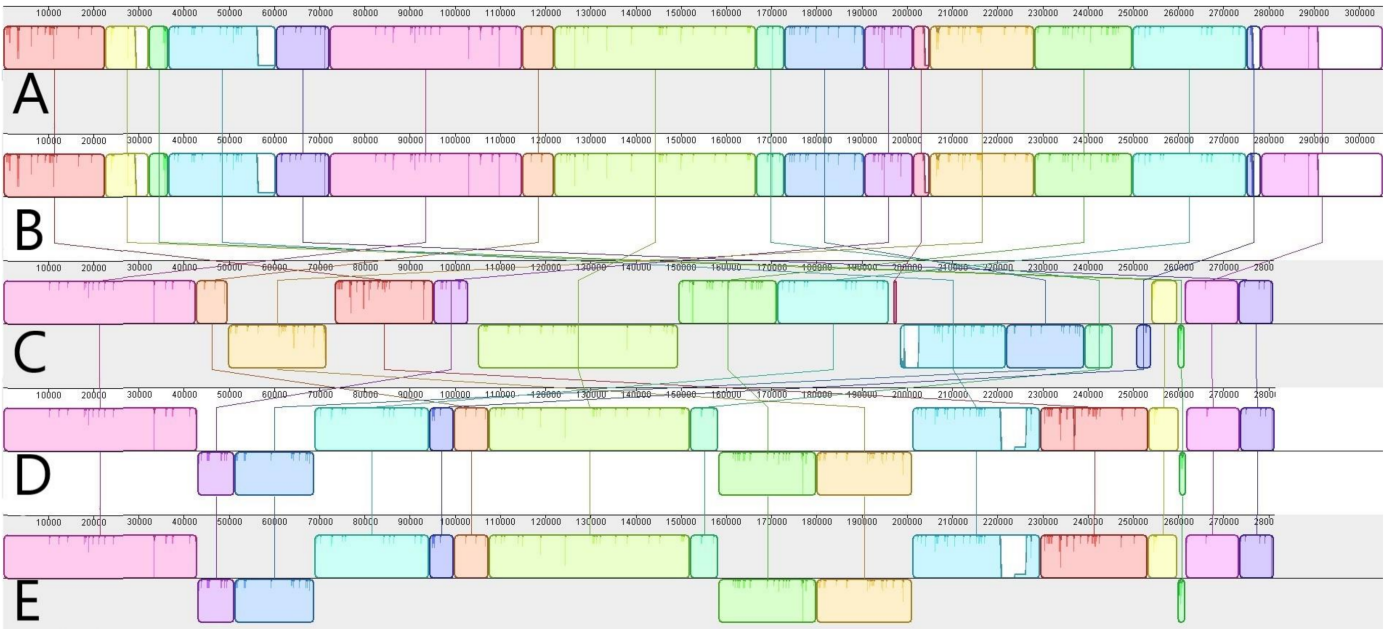

Figure 1. The mitochondrial genomes alignment of sunflower hybrids and their parental forms: (**A**) maternal *H. annuus* with PET CMS (GenBank ID MG735191.1); (**B**) hybrid *H. annuus* (VIR117A) × *H. strumosus* (440679) hybrid; (**C**) paternal *H. strumosus* (GenBank ID MT588181.1); (**D**) hybrid *H. annuus* (VIR129A) × *H. occidentalis* (441062); and (**E**) paternal *H. occidentalis* (GenBank ID MZ147621.1).

The mitogenome of the *H. annuus* (VIR117A) × *H. strumosus* (440679) hybrid is identical to the maternal hybrid. Thus, we can speak about the maternal type of mitochondrial genome inheritance in this hybrid combination. On the other hand, the *H. annuus* (VIR129A) × *H. occidentalis* (441062) mitochondrial genome is mostly (~99%)

similar to the paternal species (*H. occidentalis*), so the paternal type of mitochondrial genome inheritance is notable.

The paternal type of mitogenome inheritance is not typical for plants [13], but it was detected in some species [14,15]. Notably, the inheritance pattern shifting from maternal to paternal due to hybridization, as recently described in cucumbers [16]. In the case of the *H. annuus* (VIR129A) × *H. occidentalis* (440679) hybrid, the mitochondrial DNA (mtDNA) exhibits some differences from the paternal. The most significant one is 208 bp insertion in the case of the hybrid's mitogenome. In addition to the insertion, we localized several variant sites (INDELS/SNPs), which are displayed in Table 1.

Table 1. Variant sites localized in the *H. annuus* (VIR129A) × *H. occidentalis* (440679) hybrid in comparison with its paternal form (*H. occidentalis*).

Type	Position in *H. occidentalis* mtDNA (MZ147621.1)	Sequence in *H. occidentalis* mtDNA	Sequence in *H. annuus* × *H. occidentalis* mtDNA
SNP	27,427	T	G
SNP	31,404	T	G
INDEL	33,285–33,306	CTTTTTTTTTATTTTTTTTTT	C
SNP	48,227	C	A
INDEL	53,448	T	TA
SNP	59,286	T	C
INDEL	60,317–60,319	AGC	A
INDEL	62,077–62,078	CT	C
INDEL	71,573	G	GA
SNP	74,422	T	G
SNP	88,201	A	G
SNP	110,132	T	G
SNP	119,931	A	C
SNP	120,778	T	C
INDEL	124,360–124,366	TAAGCC	T
SNP	124,699	T	C
SNP	130,517	G	T
SNP	145,262	C	A
SNP	163,888	G	T
SNP	172,603	G	T
SNP	173,171	C	A
SNP	180,172	T	C
SNP	181,893	G	T
SNP	196,897	C	A
SNP	199,009	C	A
INDEL	206,816–206,824	AAAAAAAC	A
INDEL	213,693	T	TC
SNP	215,865	C	A
SNP	230,955	T	G
SNP	231,470	G	T
SNP	231,537	C	A
SNP	240,885	C	A
SNP	266,122	C	A

The results point out that the hybrids' mitochondrial genomes have no rearrangements. Thus, despite a significant difference in the nuclear genomes of parental species [6,7], in the case of their hybridization, it is most likely that the circuits of regulation of mitochondrial DNA recombination [17,18] have retained their functional state.

3. Methods

3.1. Plant Material

Six sunflower hybrids (F1) were obtained between domesticated sunflower (*H. annuus*) lines with cytoplasmic male sterility phenotype and wild forms of sunflowers, including annual (*H. argophyllus*, *H. praecox*) and perennial (*H. occidentalis*, *H. strumosus*) species. The following hybrids were used in the current study: *H. annuus* (VIR100A) × *H. argophyllus* (1000), *H. annuus* (VIR114A) × *H. argophyllus* (1000), *H. annuus* (VIR100A) × *H. praecox* (560,400), *H. annuus* (VIR129A) × *H. occidentalis* (441062), *H. annuus* (HA89PET1) × *H. occidentalis* (441062), and *H. annuus* (VIR117A) × *H. strumosus* (440679). All the hybrids were obtained from the

genetic collection of the N. I. Vavilov All-Russian Institute of Plant Genetic Resources (Saint Petersburg, Russia). For DNA isolation, plant leaves (at budding stage) were used. The DNA extraction was performed with the PhytoSorb kit (Syntol, Moscow, Russia), according to the manufacturer's protocol.

3.2. NGS

Then, NGS libraries were prepared with the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA), following the manufacturer's guidelines and using 10 PCR cycles. The fragment length distribution of the prepared libraries was determined with Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA), and the concentrations were evaluated with a Qubit 4 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and qPCR. The NGS libraries were diluted to 10 pM and then sequenced on MiSeq (Illumina, San Diego, CA, USA) with MiSeq Reagent Kit v2 (500 cycles) by several independent launches and with NextSeq 500 (Illumina, San Diego, CA, USA) with a Mid Output Kit v2.5 (150 cycles). We generated more than 20.6 million 250 bp paired reads and 9.25 million 76 bp paired reads for the NGS libraries (deposited to SRA under BioProject ID PRJNA929972).

3.3. Mitochondrial Genome Assembly

Quality control of reads was provided with FastQC v0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/, accessed on 23 March 2023). We used Trimmomatic v0.39 software [19] to trim adapters and discard short or low-quality reads. Contigs were generated based on MiSeq reads (250 + 250 bp) with SPAdes Genome Assembler v3.13.1 [20] using 127 k-mer length. The whole mitochondrial genome assemblies were based on high-coverage (>100 depth) contigs, selected using the Bandage v0.8.1 [21] program for visualizing de novo assembly graphs. The genome assemblies were validated by remapping reads with Bowtie 2 v2.3.5.1. SNP calling was performed with GATK software v 4.1.4.1 (https://gatk.broadinstitute.org, accessed on 23 March 2023). Complete mitochondrial genomes were aligned with Mauve tool v2.4.0 [22].

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/data8040067/s1, Figure S1: Quality scores of *H. annuus* (VIR114A) × *H. argophillus* raw reads.

Author Contributions: Conceptualization, M.S.M. and V.A.G.; methodology, M.S.M.; software, M.S.M.; validation, M.S.M.; investigation, M.S.M.; resources, V.A.G.; data curation, V.A.G.; writing—original draft preparation, M.S.M. and V.A.G.; writing—review and editing, M.S.M. and V.A.G.; visualization, M.S.M.; supervision, V.A.G.; project administration, M.S.M.; funding acquisition, M.S.M. All authors have read and agreed to the published version of the manuscript.

Funding: The reported study was funded by Russian Foundation for Basic Research (RFBR), project number 19-34-60006. The NextSeq sequencing was performed with the support of the Institute for Information Transmission Problems (Laboratory of Plant Genomics), project # FFNU-2022-0037.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available as BioProject PRJNA929972 in National Center for Biotechnology Information database.

Acknowledgments: We thank the reviewers and the editor for their suggestions and comments on the paper, which have helped us to improve the manuscript. We are also grateful to The Systems Biology Program of Skoltech, which is held with a support from Philip Morris International.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Katche, E.; Quezada-Martinez, D.; Katche, E.I.; Vasquez-Teuber, P.; Mason, A.S. Interspecific Hybridization for *Brassica* Crop Improvement. *Crop. Breed. Genet. Genom.* **2019**, *1*, 190007. [CrossRef]

2. Arriola, P.E.; Ellstrand, N.C. Crop-to-Weed Gene Flow in the Genus *Sorghum* (Poaceae): Spontaneous Interspecific Hybridization between Johnsongrass, *Sorghum halepense*, and Crop Sorghum, *S. bicolor*. *Am. J. Bot.* **1996**, *83*, 1153–1159. [CrossRef]
3. Christov, M. Helianthus Species in Breeding Research on Sunflower. In Proceedings of the 17th International Sunflower Conference, Cordoba, Spain, 8–12 June 2008.
4. Feng, J.; Jan, C.-C. Introgression and Molecular Tagging of Rf (4), a New Male Fertility Restoration Gene from Wild Sunflower *Helianthus maximiliani* L. *Theor. Appl. Genet.* **2008**, *117*, 241–249. [CrossRef] [PubMed]
5. Goryunov, D.V.; Anisimova, I.N.; Gavrilova, V.A.; Chernova, A.I.; Sotnikova, E.A.; Martynova, E.U.; Boldyrev, S.V.; Ayupova, A.F.; Gubaev, R.F.; Mazin, P.V.; et al. Association Mapping of Fertility Restorer Gene for CMS PET1 in Sunflower. *Agronomy* **2019**, *9*, 49. [CrossRef]
6. Ostevik, K.L.; Samuk, K.; Rieseberg, L.H. Ancestral Reconstruction of Karyotypes Reveals an Exceptional Rate of Nonrandom Chromosomal Evolution in Sunflower. *Genetics* **2020**, *214*, 1031–1045. [CrossRef] [PubMed]
7. Todesco, M.; Owens, G.L.; Bercovich, N.; Légaré, J.-S.; Soudi, S.; Burge, D.O.; Huang, K.; Ostevik, K.L.; Drummond, E.B.M.; Imerovski, I.; et al. Massive Haplotypes Underlie Ecotypic Differentiation in Sunflowers. *Nature* **2020**, *584*, 602–607. [CrossRef] [PubMed]
8. Sukno, S.; Ruso, J.; Jan, C.C.; Melero-Vara, J.M.; Fernández-martínez, J.M. Interspecific Hybridization between Sunflower and Wild Perennial Helianthus Species via Embryo Rescue. *Euphytica* **1999**, *106*, 69–78. [CrossRef]
9. Hristova-Cherbadzi, M. Characterization of Hybrids, Forms and Lines, Obtained from Interspecific Hybridization of Cultivated Sunflower *Helianthus annuus* L. with Wild Species of Genus Helianthus. *Biotechnol. Biotechnol. Equip.* **2009**, *23*, 112–116. [CrossRef]
10. Makarenko, M.S.; Kornienko, I.V.; Azarin, K.V.; Usatov, A.V.; Logacheva, M.D.; Markin, N.V.; Gavrilova, V.A. Mitochondrial Genomes Organization in Alloplasmic Lines of Sunflower (*Helianthus annuus* L.) with Various Types of Cytoplasmic Male Sterility. *PeerJ* **2018**, *6*, e5266. [CrossRef] [PubMed]
11. Makarenko, M.S.; Omelchenko, D.O.; Usatov, A.V.; Gavrilova, V.A. The Insights into Mitochondrial Genomes of Sunflowers. *Plants* **2021**, *10*, 1774. [CrossRef] [PubMed]
12. Makarenko, M.; Usatov, A.; Tatarinova, T.; Azarin, K.; Kovalevich, A.; Gavrilova, V.; Horn, R. The Investigation of Perennial Sunflower Species (*Helianthus* L.) Mitochondrial Genomes. *Genes* **2020**, *11*, 982. [CrossRef] [PubMed]
13. Knoop, V.; Volkmar, U.; Hecht, J.; Grewe, F. Mitochondrial Genome Evolution in the Plant Lineage. In *Plant Mitochondria*; Kempken, F., Ed.; Advances in Plant Biology; Springer: New York, NY, USA, 2011; pp. 3–29. ISBN 978-0-387-89781-3.
14. Chat, J.; Chalak, L.; Petit, R.J. Strict Paternal Inheritance of Chloroplast DNA and Maternal Inheritance of Mitochondrial DNA in Intraspecific Crosses of Kiwifruit. *Theor. Appl. Genet.* **1999**, *99*, 314–322. [CrossRef]
15. Fauré, S.; Noyer, J.-L.; Carreel, F.; Horry, J.-P.; Bakry, F.; Lanaud, C. Maternal Inheritance of Chloroplast Genome and Paternal Inheritance of Mitochondrial Genome in Bananas (*Musa acuminata*). *Curr. Genet.* **1994**, *25*, 265–269. [CrossRef] [PubMed]
16. Park, H.-S.; Lee, W.K.; Lee, S.-C.; Lee, H.O.; Joh, H.J.; Park, J.Y.; Kim, S.; Song, K.; Yang, T.-J. Inheritance of Chloroplast and Mitochondrial Genomes in Cucumber Revealed by Four Reciprocal F1 Hybrid Combinations. *Sci. Rep.* **2021**, *11*, 2506. [CrossRef] [PubMed]
17. Gualberto, J.M.; Mileshina, D.; Wallet, C.; Niazi, A.K.; Weber-Lotfi, F.; Dietrich, A. The Plant Mitochondrial Genome: Dynamics and Maintenance. *Biochimie* **2014**, *100*, 107–120. [CrossRef] [PubMed]
18. Morley, S.A.; Nielsen, B.L. Plant Mitochondrial DNA. *Front. Biosci. Landmark* **2017**, *22*, 1023–1032. [CrossRef]
19. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]
20. Nurk, S.; Bankevich, A.; Antipov, D.; Gurevich, A.; Korobeynikov, A.; Lapidus, A.; Prjibelsky, A.; Pyshkin, A.; Sirotkin, A.; Sirotkin, Y.; et al. Assembling Genomes and Mini-Metagenomes from Highly Chimeric Reads. In Proceedings of the Research in Computational Molecular Biology; Deng, M., Jiang, R., Sun, F., Zhang, X., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 158–170.
21. Wick, R.R.; Schultz, M.B.; Zobel, J.; Holt, K.E. Bandage: Interactive Visualization of de Novo Genome Assemblies. *Bioinformatics* **2015**, *31*, 3350–3352. [CrossRef] [PubMed]
22. Darling, A.C.E.; Mau, B.; Blattner, F.R.; Perna, N.T. Mauve: Multiple Alignment of Conserved Genomic Sequence with Rearrangements. *Genome Res.* **2004**, *14*, 1394–1403. [CrossRef] [PubMed]