

Supplemental Materials

1. Materials and Methods

1.1 Primer design

An advantage to the NCBI Primer-BLAST is that it can accept input of multiple sequences in a single fasta file, and if possible, generate primers usable with all the input sequences.

1.2 PCR protocol

PCR was performed using Phusion DNA polymerase (Fisher Scientific), with the standard HF buffer, using a hot start protocol: a 30 second precycle at 98°C, followed by 40 cycles (30 seconds melting at 98°C, 30 seconds annealing at the defined temperature, 30 seconds extension at 72°C), followed by 10 minutes at 72°C then stopping at 4°C. For all the primers employed in this study, annealing was successful at either 57°C. PCR was assayed on 1% Tris-acetate agarose gels with SafeRed to visualize DNA.

1.3 Genome sequencing

Sequencing was performed at the Centre d'Expertise et Services of Genome Quebec. 2x151bp paired-end sequencing was done on an Illumina NovaSeq 6000 instrument. 320,869,080 total paired reads were generated, totalling 96.9 Gbp of raw sequence, or 69-fold coverage assuming a genome size of 1.4 Gbp. Per FastP, the average insert size of the input library was estimated to be 218bp, hence there was overlap between the forward and reverse read fastq files for many inserts.

1.4 Informatics

1.4.1 Sanger sequencing

Sanger sequence traces from the Genome Quebec core laboratory were analyzed with CodonCode Aligner (<https://www.codoncode.com/aligner/>). After loading the traces, bases were called and sequences were trimmed using the suite tools, followed by contig assembly. As needed, sequences of the relevant genes from the relevant species were directly loaded as fasta files, and included in the contig alignments. Sequences of genes from various *Chenopodium* species were obtained from the NCBI Nucleotide database, or from the iBOLD database (<https://www.boldsystems.org/>) maintained at the University of Guelph in Ontario, Canada.

1.4.2 Raw sequence quality control, cleanup, and genome size estimate by k-mer analysis

For analysis of the whole genome sequences, raw data quality was assessed using FastQC [1]. K-mer analysis was performed using KMC3 [2]. We compared this to Jellyfish, and the two tools generated similar results but KMC3 was significantly easier to implement, with the option to run it locally on a Windows OS system. KMC3 output was modelled using the online GenomeScope server [3]; models could be generated using k25 and k50 counts, with similar results, but GenomeScope was unable to generate a model using a k100 count.

The raw Illumina fastq files were cleaned using FastP [4], with parameters restricted to removing adapter sequences and polyG sequences, the latter being a known systematic issue with NovaSeq chemistry. The documentation for Platanus recommends not performing pre-assembly base quality

trimming, as quality trims are intrinsic to the Platanus assembler, which works best with the most complete input sequence files especially with paired end data files.

1.4.3 Whole genome assembly

Genome assembly was performed using both Platanus [5] and Abyss [6], trying a variety of parameters including different initial k-mers and allowed mismatch percentages. We used Platanus v1. We did not attempt to use Platanus v2 (allee) as we expected minimal heterozygosity and it is unclear how Platanus-allee deals with polyploid genomes versus haplotypes or diploids. Statistics of the assemblies were compared using scripts in the BBTools suite (<https://sourceforge.net/projects/bbmap>), and with QUAST [7]. Note that the BBTools output reverses the N50/L50 statistics with respect to standard nomenclature.

1.4.4 Repeat sequence analysis

Repeat sequences were assessed using RepeatMasker version 4.1.1 [8], run with rmbblastn version 2.10.0+; query was compared to a set of quinoa genome repeat elements annotated in the PlantRep database [9].

1.4.5 BUSCO analysis of standardized gene set completion

Genome completion was assessed using BUSCO (Benchmarking Universal Single-Copy Orthologue) v5.2.2 [10], using the eudicots_odb10 database, based on 31 genomes, as a reference.

1.4.6 Variant calling

Variant calling was performed by aligning the raw fastq reads to the Platanus k96 genome assembly previously generated from those reads. The goal was to assess the degree of heterozygosity, not to call variants with respect to other published genomes. First, sambamba was used to remove duplicate reads, sort and index the raw fastq files [11]. Second, bwa-mem was run with mismatch parameter 6 (default is 4) [12], followed by Bcftools-mpileup with a maximum read depth of 400 to reduce running time (although oddly in the resulting variant vcf files there are still a small number of variants with greater than 400 read depth). Bcftools call was run to call variants, using default stringency parameters [13,14]. For statistics of variant calling the genome assembly, insertion/deletion variants were excluded, and contigs shorter than 300bp were excluded (these are mostly repetitive sequences in any case). To further restrict variants to non-repetitive sequences, BEDOPs was used to convert the RepeatMasker output file to bed format [15]. BEDTools was used to filter the variant vcf file to exclude sites in the RepeatMasker bed file [16]. For variant calling, the genomes of *C. quinoa* (version 2) and *C. berlandieri nuttaliae* were downloaded in fasta format from the CoGe database. We used vcflib stats to obtain variant genotype statistics; bcftools stats gave slightly different numbers for total, heterozygous and homozygous SNPs for unclear reasons, but the differences do not affect our conclusions.

1.4.7 Dot plot comparisons of organelle genomes

Dot plots of genome assembly scaffolds to *C. quinoa* chloroplast and mitochondrial genomes were performed using the online D-GENIES server [17].

1.4.8 Individual gene model building, intra- and inter-species comparisons

For sequence comparison of individual genes (e.g. *SOS1*), reference sequences and spliced gene models were obtained for the genes in question from the *C. pallidicaule*, *C. suecicum*, *C. quinoa* v2 and *C. berlandieri nuttaliae* annotated genomes in the Comparative Genomics (CoGe) database (<https://genomevolution.org/coge/>) In some cases, gene models were revised using the spliced model from another species with the genomic sequence of the given species. This improved consistency of gene comparisons across the species, while making the description of the various genes in the Results section somewhat hard to describe. Sources for finding the genes varied, sometimes *Arabidopsis thaliana*, sometimes *C. quinoa* itself. One or more of these reference genes were used to blast our assembled SV89-10 Canadian *C. berlandieri* genome, at either the DNA or protein level, using the Platanus k-mer 96 gap_closed assembly. In some cases, genes were fragmented across multiple scaffolds in our assembly. In such cases, scaffolds containing equivalent coding exons from 5' to 3' were manually combined. Given the two parental sub-genomes of *C. berlandieri*, it was important not to mis-assemble scaffolds containing different exons from each sub-genome; fortunately, in all cases that we examined, individual assembled scaffolds containing protein-coding exons were more similar to either the A or B sub-genome of the reference genomes. In all genes examined, there were no discrepancies between assignment of the A and B homologs between the two tetraploid references (*C. quinoa* v2, *C. berlandieri nuttaliae*). Spliced gene models were built using either the A or B sub-genome homolog from the *C. berlandieri nuttaliae* annotated genes with the manually assembled A or B sub-genome scaffolds of our assembly; if the *C. berlandieri nuttaliae* gene modelling was ambiguous, some of the other reference species were used alternatively. Splign, SIM4 and GMAP were all used for model building, with essentially identical results [18-20].

Minimum scaffold length	Number of scaffolds	Number of contigs	Total scaffold length	Total contig length	Scaffold contig coverage
All	2,223,105	2,228,384	1,308,412,757	1,308,135,794	99.98%
100	2,223,105	2,228,384	1,308,412,757	1,308,135,794	99.98%
250	790,154	795,430	1,091,061,749	1,090,800,272	99.98%
500	351,479	356,753	941,537,391	941,276,024	99.97%
1 KB	176,632	181,906	821,399,569	821,138,202	99.97%
2.5 KB	105,493	110,633	701,949,973	701,694,206	99.96%
5 KB	54,911	59,024	521,340,195	521,133,889	99.96%
10 KB	17,506	19,614	259,613,574	259,507,514	99.96%
25 KB	843	1,074	25,535,499	25,524,026	99.96%
50 KB	9	19	488,766	488,248	99.89%

Table S1. Statistics of Abyss genome assembly with k-mers=104. Main genome scaffold N/L50: 88,430/3.125 KB. Main genome contig N/L50: 91,956/3.02 KB. Main genome scaffold N/L90: 1,218,602/207bp. Main genome contig N/L90: 1,208,431/207bp. Max scaffold length: 76.235 KB. Max contig length: 66.776 KB. Number of scaffolds > 50 KB: 9. Percent main genome in scaffolds > 50 KB: 0.04%. GC content: 0.37.

Number of genes	Gene status in genome assembly
1486	Complete BUSCOs (C)
737	Complete and single-copy BUSCOs (S)
749	Complete and duplicated BUSCOs (D)
426	Fragmented BUSCOs (F)
414	Missing BUSCOs (M)
2326	Total BUSCO groups searched

Table S2. BUSCO analysis of Abyss k104 assembly using a database of eudicot-specific genes. # BUSCO version is: 5.2.2. # The lineage dataset is: eudicots_odb10 (Creation date: 2020-09-10, number of genomes: 31, number of BUSCOs: 2326). # Gene predictor used: metaeuk

Figures

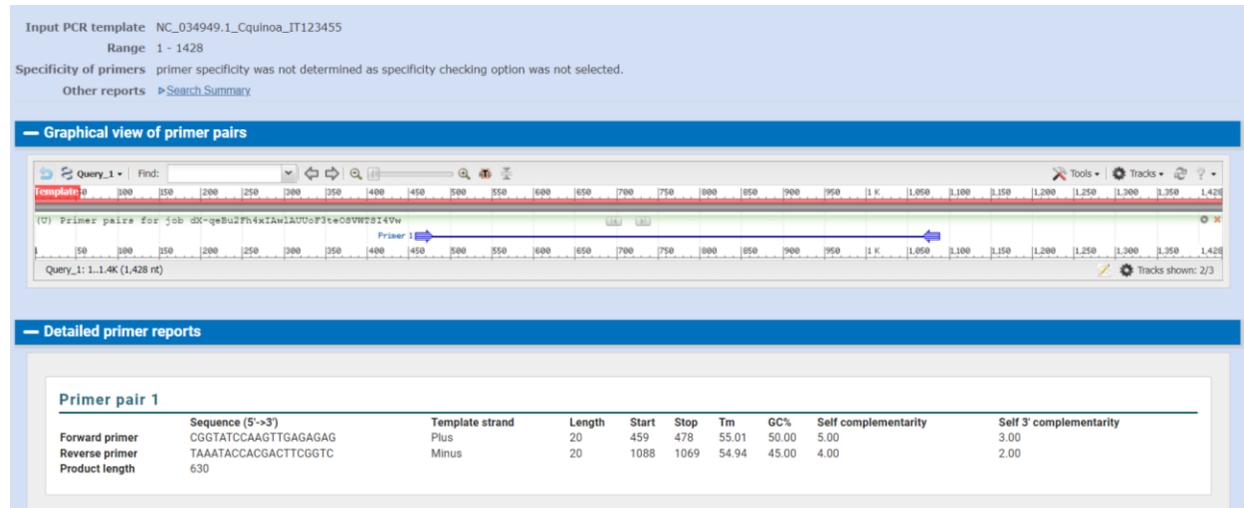
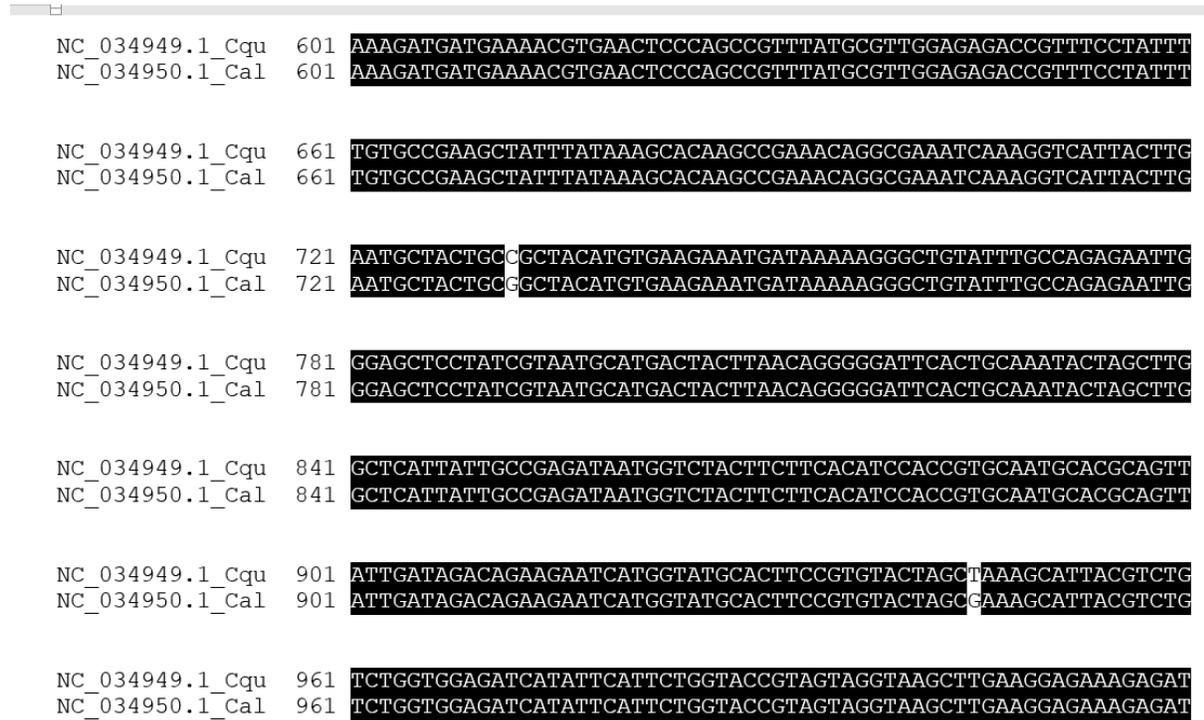


Figure S1. Primer design of *rbcl*. a, alignment of *rbcl* in *C. quinoa* and *C. album* gene as amplified by primers *rbcl*M1-F and *rbcl*M1-R, showing two sites different in the amplicon. b, PCR amplification product of *C. quinoa rbcl* using primers *rbcl*M1-F and *rbcl*M1-R, per NCBI Primer-BLAST

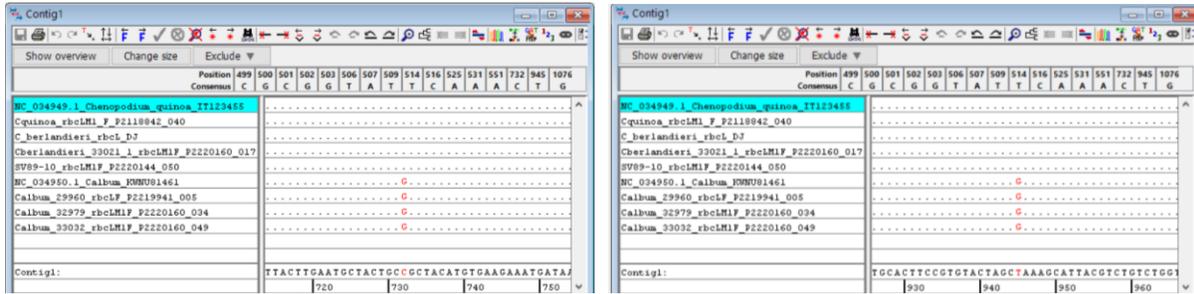


Figure S2. *rbcL* barcoding. Alignment of references and PCR-amplified DNAs, using primers *rbcLM1-F* and *rbcLM2-R*, sequenced in the forward direction using *rbcLM1-F* (identical sequences on reverse strands). Numbering is from the ATG initiation codon of the *rbcL* gene in the three references. *C. quinoa* and *C. album* reference sequences from NCBI are given with NC_ numbers, *C. berlandieri* reference sequence is from chloroplast genome unpublished, courtesy of David Jarvis. Positive control DNAs of *C. quinoa*, *album* and *berlandieri* were obtained from leaves grown from seeds obtained from David Brenner at the USDA as described in Materials and Methods. Plant SV89-10, collected in Manitoba, Canada, was obtained from Sara Halwas and Anne Worley as described in Materials and Methods. The two sites in *rbcL* different between *C. quinoa* and *C. album* in the amplified region are included.

Samuels Genomic sequence of Canadian *Chenopodium berlandieri*

~~satK_Calypso~~ 361 AAGAAAAAAGAAATAGTAAATCTCATAATTTGCGATCTATTCATTCAATATTTCTT
~~satK_Cuinao~~ 359 -----AAAAAAGAAATAGTAAATCTCATAATTTGCGATCTATTCATTCAATATTTCTT

~~satK_Calypso~~ 421 TTTTAGAGGACAAATTCCTCCATTCAATTATGTGTTAGATATATTAATAACTTACCCT
~~satK_Cuinao~~ 412 TTTTAGAGGACAAATTCCTCCATTCAATTATGTGTTAGATATATTAATAACTTACCCT

~~satK_Calypso~~ 481 GCCCATCTAGAAATCTTGTTCAAACCTCTCGCTACTGGTTGAAAGATGCCTCTTCTTTG
~~satK_Cuinao~~ 472 GCCCATCTAGAAATCTTGTTCAAACCTCTCGCTACTGGTTGAAAGATGCCTCTTCTTTG

~~satK_Calypso~~ 541 CATTAGTAAGATTTTTCTTTATGAGTCTCGTAATTTAAATAGTCTTATTATCCAAAG
~~satK_Cuinao~~ 532 CATTAGTAAGATTTTTCTTTATGAGTCTCGTAATTTAAATAGTCTTATTATCCAAAG

~~satK_Calypso~~ 601 GAATTCATTTCTTTTTGAAAAAAGGAATCAAAGATTATTCTTGTTCCTATATAAATTC
~~satK_Cuinao~~ 592 GAATTCATTTCTTTTTGAAAAAAGGAATCAAAGATTATTCTTGTTCCTATATAAATTC

~~satK_Calypso~~ 661 CATGTATCGAATACGAATCTTTTTGTTTTTCTCCGCAACCAATCCTCTATTACGA
~~satK_Cuinao~~ 652 CATGTATCGAATACGAATCTTTTTGTTTTTCTCCGCAACCAATCCTCTATTACGA

~~satK_Calypso~~ 721 TCACCTCTTTTGAAACCCCTTATTGAACGAATTTTTCTACGGAAAACAAAATATCTA
~~satK_Cuinao~~ 712 TCACCTCTTTTGAAACCCCTTATTGAACGAATTTTTCTACGGAAAACAAAATATCTA

~~satK_Calypso~~ 781 GTAAAAGTTTTACTAAGGATTTGGGGTTATCCTATGGCTTTTCAGAGAACCCTTCCCG
~~satK_Cuinao~~ 772 GTAAAAGTTTTACTAAGGATTTGGGGTTATCCTATGGCTTTTCAGAGAACCCTTCCCG

~~satK_Calypso~~ 841 CATTATGTTAGGTATCAAGGAAATCAATTCTAGCCTCAAAGGGACATCTCTTTTGATG
~~satK_Cuinao~~ 832 CATTATGTTAGGTATCAAGGAAATCAATTCTAGCCTCAAAGGGACATCTCTTTTGATG

~~satK_Calypso~~ 901 CATAAATGGAAAAATTACCTTATCCATTCTGGCAATGTAATTTTCTGTGGTCTCAA
~~satK_Cuinao~~ 892 CATAAATGGAAAAATTACCTTATCCATTCTGGCAATGTAATTTTCTGTGGTCTCAA

~~satK_Calypso~~ 961 CCAAAAAGCTCTATATCAATCGATTATCAAACCATTCCTTAGACTTTATGGGTTTTCTT
~~satK_Cuinao~~ 952 CCAAAAAGCTCTATATCAATCGATTATCAAACCATTCCTTAGACTTTATGGGTTTTCTT

~~satK_Calypso~~ 1021 TCAAATGTGCGACTCAATTCTTCAGTAATACGGAGTCAATGTTGAAAAATGCATTTCTA
~~satK_Cuinao~~ 1012 TCAAATGTGCGACTCAATTCTTCAGTAATACGGAGTCAATGTTGAAAAATGCATTTCTA

~~satK_Calypso~~ 1081 ATAGAGAATATTATTAAGAAATTCGATGCCACAGTCCAAATTAATTCCTCTGGTTGGATCG
~~satK_Cuinao~~ 1072 ATAGAGAATATTATTAAGAAATTCGATGCCACAGTCCAAATTAATTCCTCTGGTTGGATCG

~~satK_Calypso~~ 1141 TTGGCTAAAGCGAAATTTGTAATGTATTAGGGCATCCCCTTAGTAAGTCGGTCTGGACC
~~satK_Cuinao~~ 1132 TTGGCTAAAGCGAAATTTGTAATGTATTAGGGCATCCCCTTAGTAAGTCGGTCTGGACC

~~satK_Calypso~~ 1201 GATTTATCCGATTCTGATATTATTGATCGATTGGACGATATGTAGAATATTTTTCAT
~~satK_Cuinao~~ 1192 GATTTATCCGATTCTGATATTATTGATCGATTGGACGATATGTAGAATATTTTTCAT

~~satK_Calypso~~ 1261 TATTATAGTGGCTCTTCAGAAAAAAGTTTGTATCGAATAAAGTATATACTTCGATT
~~satK_Cuinao~~ 1252 TATTATAGTGGCTCTTCAGAAAAAAGTTTGTATCGAATAAAGTATATACTTCGATT

~~satK_Calypso~~ 1321 TCTTGTGCTAGAACTTTGTCTCGTAAACACAAAAGTACTGTACGTGCTTTTTGAAAAGA
~~satK_Cuinao~~ 1312 TCTTGTGCTAGAACTTTGTCTCGTAAACACAAAAGTACTGTACGTGCTTTTTGAAAAGA



Figure S3. Primer design of matK. a, alignment of *matK* in *C. quinoa* vs *C. album* as amplified by matK390-F and matK3FKIM-R, showing six sites different in the amplicon. b, PCR amplification product of *C. quinoa matK* using primers matK390-F and matK3FK1M-R, per NCBI Primer-BLAST.

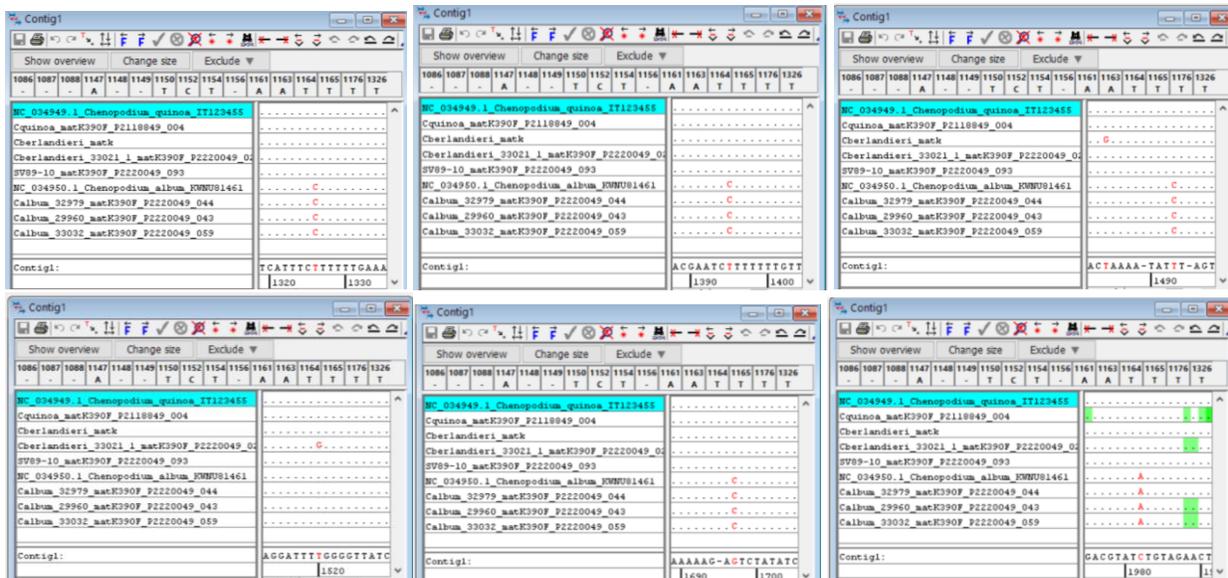


Figure S4. *matK* barcoding. Alignment of references and PCR-amplified DNAs, using primers matK390-F and matK3FKIM-R, sequenced in the forward direction using matK390-F (identical sequences on reverse strands). Numbering is from a position in the *C. berlandieri* chloroplast genome approximately 710 bp upstream of the *matK* ATG initiation codon. *C. quinoa* and *C. album* reference sequences from NCBI are given with NC_ numbers, *C. berlandieri* reference sequence is from chloroplast genome unpublished, courtesy of David Jarvis. Positive control DNAs of *C. quinoa*, *album* and *berlandieri* were obtained from leaves grown from seeds obtained from David Brenner at the USDA as described in Materials and Methods. Plant SV89-10, collected in Manitoba, Canada, was obtained from Sara Halwas and Anne Worley as described in Materials and Methods. Five sites in *matK* different between *C. quinoa* and *C. album* are shown (the sixth different site is too close to one primer for clean sequence), and one site different in the USDA *C. berlandieri* control DNA.

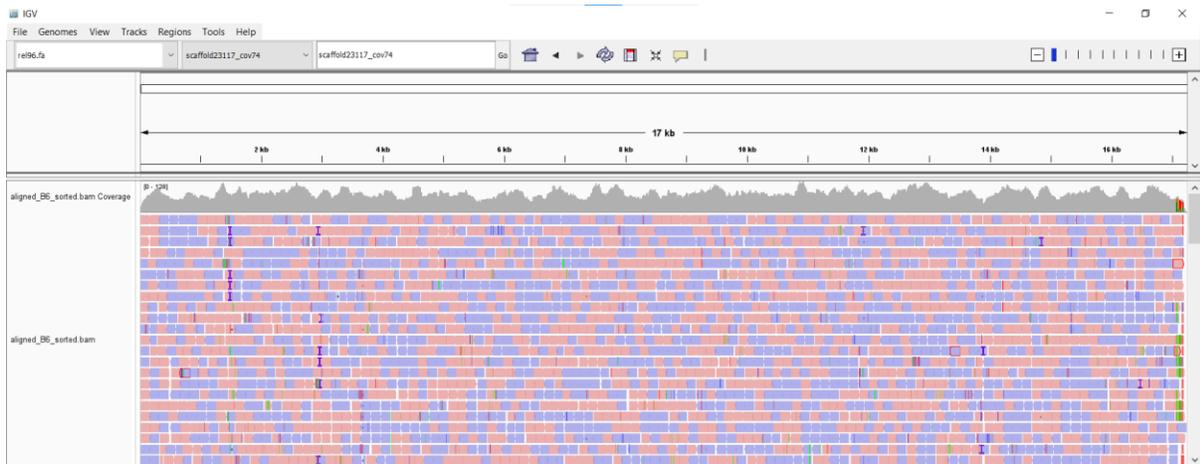
Samuels Genomic sequence of Canadian *Chenopodium berlandieri*

A.



Samuels Genomic sequence of Canadian *Chenopodium berlandieri*

B.



Samuels Genomic sequence of Canadian *Chenopodium berlandieri*

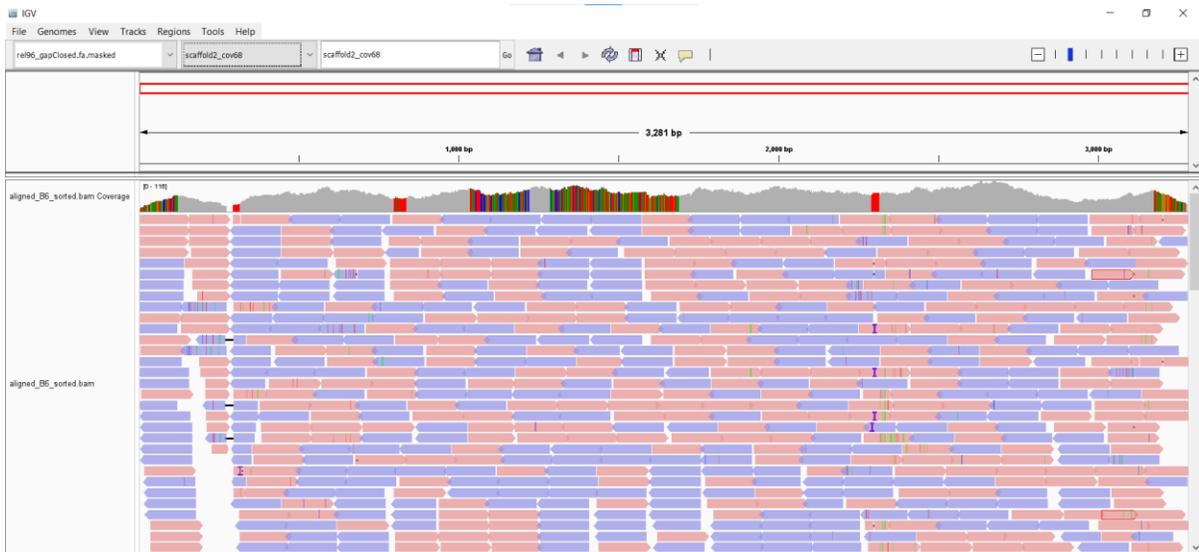
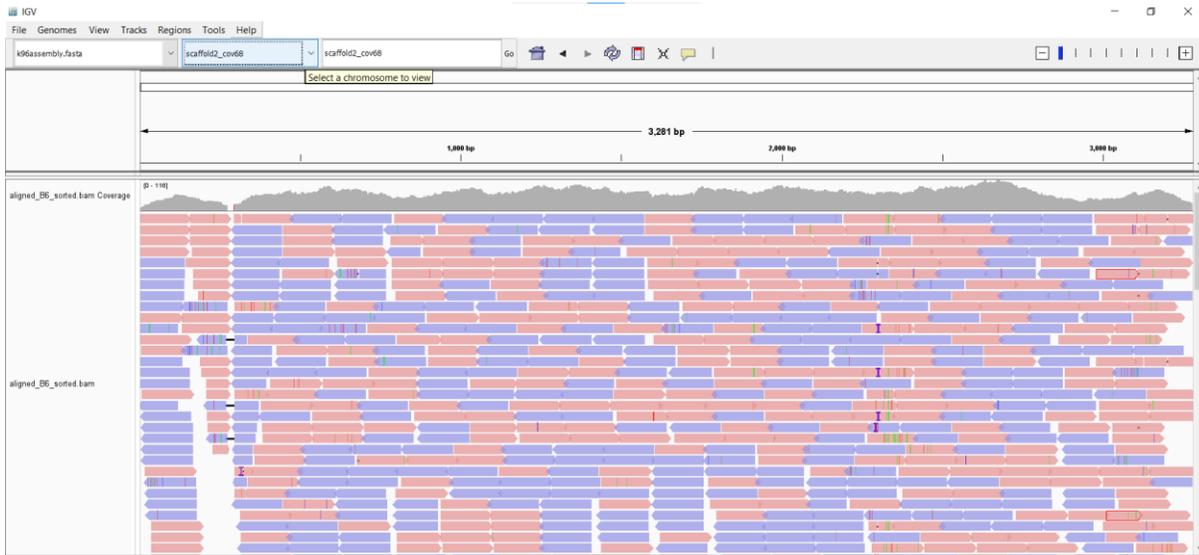
C.



D.



E.



F.

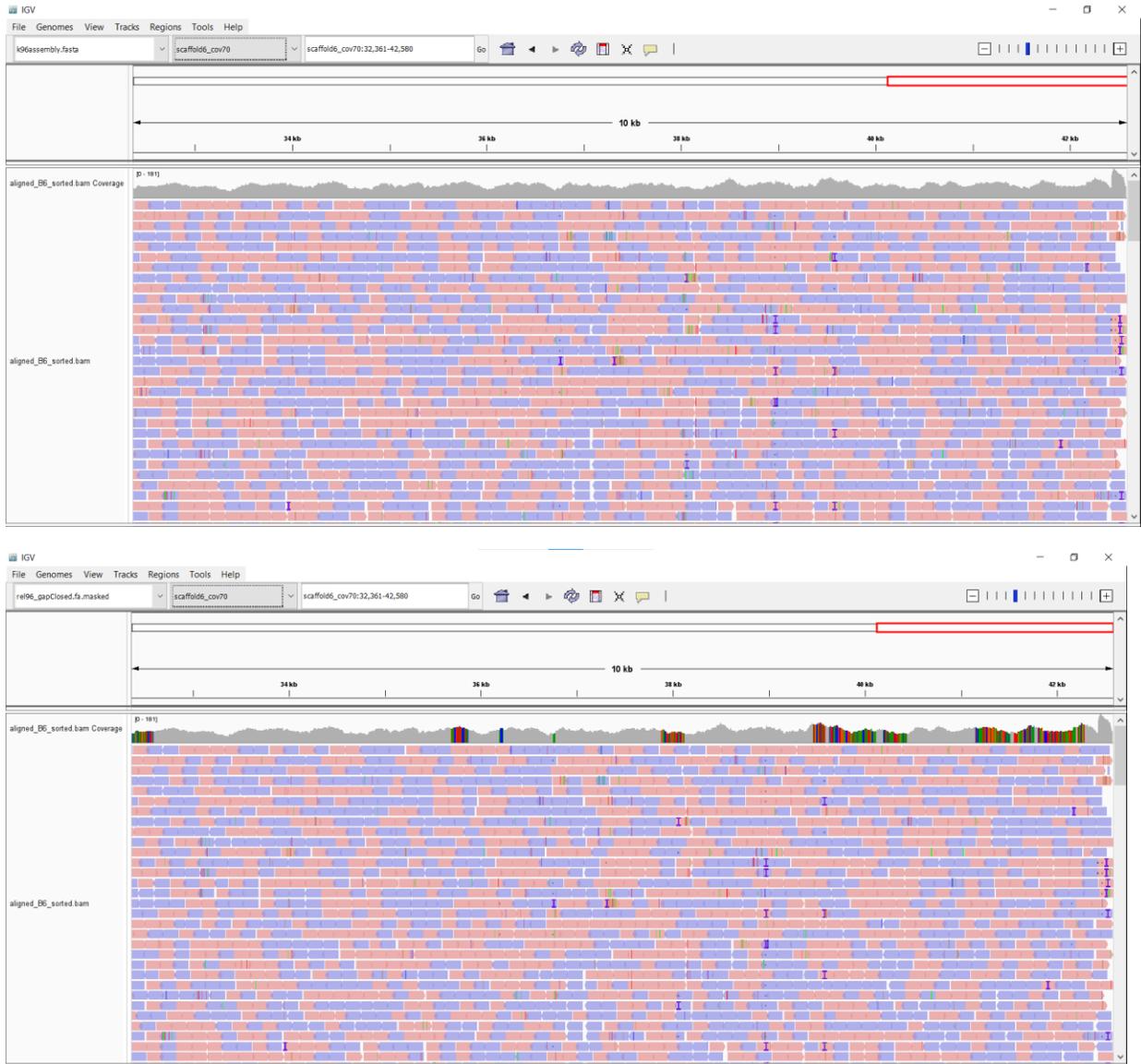
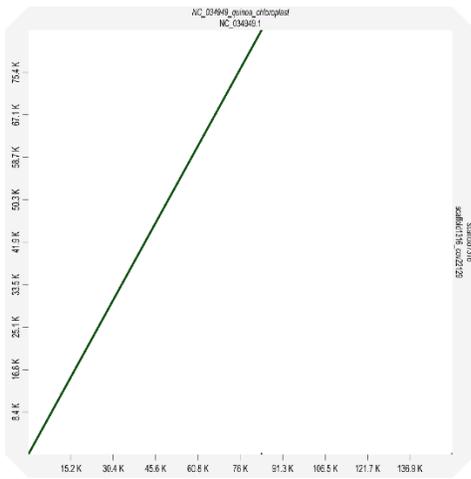
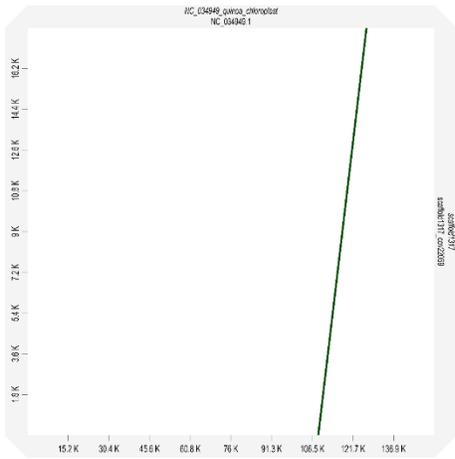


Figure S5. Selected scaffolds showing examples of heterozygous SNPs. Each scaffold is shown twice, first aligned to the whole genome assembly, then below aligned to the genome assembly with repeats masked. Split vertical colored lines present in both alignments of the scaffold represent heterozygous sites. Solid vertical colored lines present in only the lower alignment represent repetitive regions (where the bases called in the scaffold are different from the N's in the masked genome assembly, hence are shown as solid colored lines). A-C, scaffolds of *SOS1* homologs: A, scaffold33144; B, scaffold23117; C, scaffold473876. D-F, miscellaneous scaffolds: D, scaffold7603; E, scaffold2; F, scaffold6.

A.



B.



C.

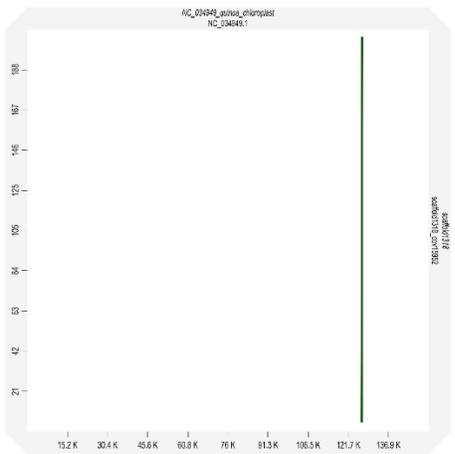
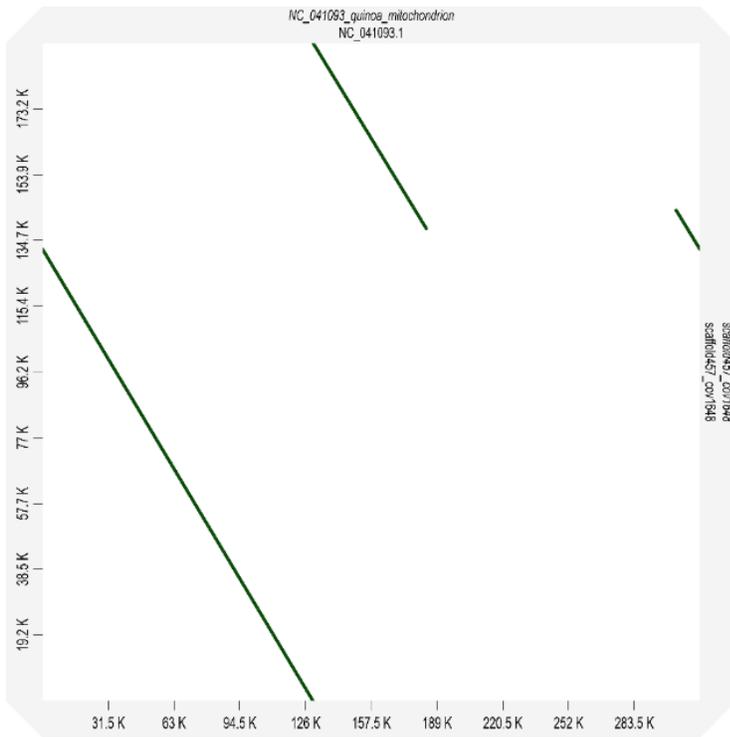


Figure S6. Chloroplast genome assembly. A, B, C, alignments to *C. quinoa* unique chloroplast genome regions.

A.



B.

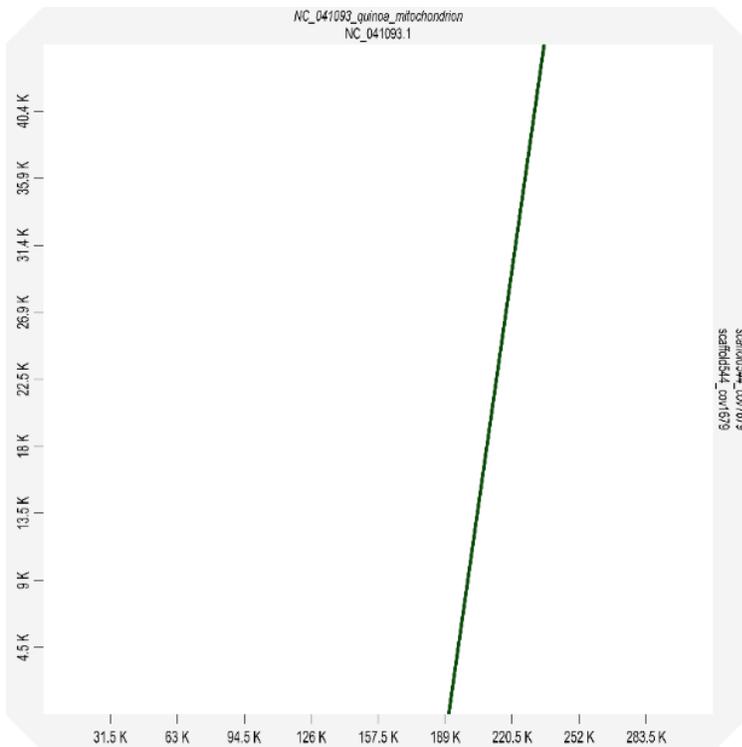


Figure S7. Mitochondrial genome assembly. A, B, alignments to *C. quinoa* mitochondrial genome.

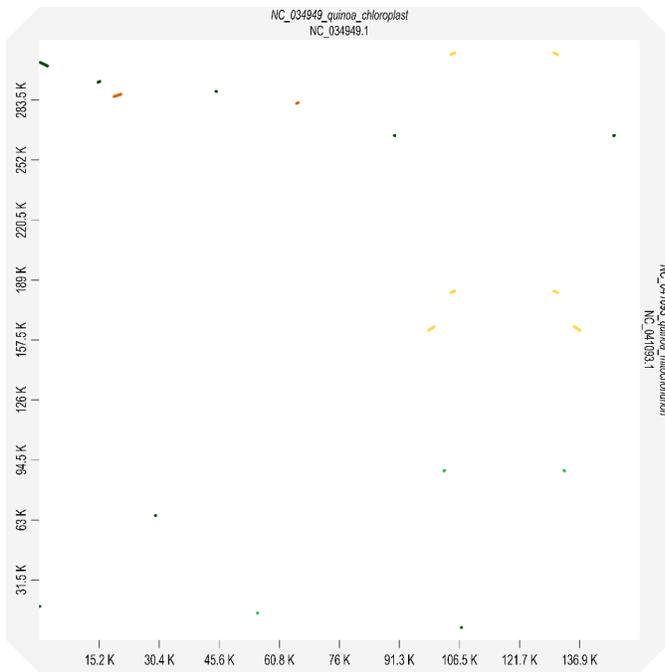


Figure S8. Alignment of *C. quinoa* chloroplast and mitochondrial genomes to each other.

Samuels Genomic sequence of Canadian *Chenopodium berlandieri*

A.

```

Cqinoa_SQS1-C6B1-1234 1 MAAYRI AVFPFLMLAEVGAETTTPTNSTI SSSAMAESEESTPTDAVIF FOVSLILO IACHRF LROTRVPTVALL IGI OLGSLDSEVIT SVSEYTKHGLR IGDDI RIWENIDPELLAVL FPALEFFESSFMEIHOIKRGAADMLL 150
Cqinoa_SQS1-1211 1 MAAYRI AVFPFLMLAEVGAETTTPTNSTI SSSAMAESEESTPTDAVIF FOVSLILO IACHRF LROTRVPTVALL IGI OLGSLDSEVIT SVSEYTKHGLR IGDDI RIWENIDPELLAVL FPALEFFESSFMEIHOIKRGAADMLL 141
Cberlandierutatae_SQS1-C6B1-1207 1 MAAYRI AVFPFLMLAEVGAETTTPTNSTI SSSAMAESEESTPTDAVIF FOVSLILO IACHRF LROTRVPTVALL IGI OLGSLDSEVIT SVSEYTKHGLR IGDDI RIWENIDPELLAVL FPALEFFESSFMEIHOIKRGAADMLL 137
SVB9-10_SQS1-C6Bhomolog1-1207 1 MAAYRI AVFPFLMLAEVGAETTTPTNSTI SSSAMAESEESTPTDAVIF FOVSLILO IACHRF LROTRVPTVALL IGI OLGSLDSEVIT SVSEYTKHGLR IGDDI RIWENIDPELLAVL FPALEFFESSFMEIHOIKRGAADMLL 137

Cqinoa_SQS1-C6B1-1234 161 SVYAVYF TAGEGADVSQVLTVMITLQMF YAAARATAF KQESQSLHFFHEWEMVAY IANTLIF IL S6AVIAGVLS50N1 FENHNASVSKLSVNSYMSMEDVNLQ I EKWIOHHTLGRSDLSVGVVLF ELFDG2ATGWVYLLLVVYLVARGV 200
Cqinoa_SQS1-1211 162 AGPVLV ISTFLCGAALK ..... LFFYDWSWKTSL LQGLLSATDPVAVALLKELGASKSLSTI IEGESLMDGDT ..... AI VVYVQLFKMILGRTFNWAS ILKYLVQVTFGAVGFI AFG IASVWLQFI FNDTVIEITLTLAV 277
Cberlandierutatae_SQS1-C6B1-1207 163 AGPVLV ISTFLCGAALK ..... LFFYDWSWKTSL LQGLLSATDPVAVALLKELGASKSLSTI IEGESLMDGDT ..... AI VVYVQLFKMILGRTFNWAS ILKYLVQVTFGAVGFI AFG IASVWLQFI FNDTVIEITLTLAV 273
SVB9-10_SQS1-C6Bhomolog1-1207 163 AGPVLV ISTFLCGAALK ..... LFFYDWSWKTSL LQGLLSATDPVAVALLKELGASKSLSTI IEGESLMDGDT ..... AI VVYVQLFKMILGRTFNWAS ILKYLVQVTFGAVGFI AFG IASVWLQFI FNDTVIEITLTLAV 273

Cqinoa_SQS1-C6B1-1234 301 SVYAVYF TAGEGADVSQVLTVMITLQMF YAAARATAF KQESQSLHFFHEWEMVAY IANTLIF IL S6AVIAGVLS50N1 FENHNASVSKLSVNSYMSMEDVNLQ I EKWIOHHTLGRSDLSVGVVLF ELFDG2ATGWVYLLLVVYLVARGV 400
Cqinoa_SQS1-1211 278 SVYAVYF TAGEGADVSQVLTVMITLQMF YAAARATAF KQESQSLHFFHEWEMVAY IANTLIF IL S6AVIAGVLS50N1 FENHNASVSKLSVNSYMSMEDVNLQ I EKWIOHHTLGRSDLSVGVVLF ELFDG2ATGWVYLLLVVYLVARGV 427
Cberlandierutatae_SQS1-C6B1-1207 274 SVYAVYF TAGEGADVSQVLTVMITLQMF YAAARATAF KQESQSLHFFHEWEMVAY IANTLIF IL S6AVIAGVLS50N1 FENHNASVSKLSVNSYMSMEDVNLQ I EKWIOHHTLGRSDLSVGVVLF ELFDG2ATGWVYLLLVVYLVARGV 403
SVB9-10_SQS1-C6Bhomolog1-1207 274 SVYAVYF TAGEGADVSQVLTVMITLQMF YAAARATAF KQESQSLHFFHEWEMVAY IANTLIF IL S6AVIAGVLS50N1 FENHNASVSKLSVNSYMSMEDVNLQ I EKWIOHHTLGRSDLSVGVVLF ELFDG2ATGWVYLLLVVYLVARGV 403

Cqinoa_SQS1-C6B1-1234 451 VVQVLYPFLCYFQYQWMEKEMAM LWAQLRQAVALSLSL SVKRS50GPAYSLTOTOTLFFVFTG I VFLFTL I I NOSTTQVLFRLGMDLKS AKARRL LEFTKYEKKALAEFDLQDEDELOPADMTVKRY I KSLNLI SGR IHPHDA 600
Cqinoa_SQS1-1211 478 VVQVLYPFLCYFQYQWMEKEMAM LWAQLRQAVALSLSL SVKRS50GPAYSLTOTOTLFFVFTG I VFLFTL I I NOSTTQVLFRLGMDLKS AKARRL LEFTKYEKKALAEFDLQDEDELOPADMTVKRY I KSLNLI SGR IHPHDA 577
Cberlandierutatae_SQS1-C6B1-1207 474 VVQVLYPFLCYFQYQWMEKEMAM LWAQLRQAVALSLSL SVKRS50GPAYSLTOTOTLFFVFTG I VFLFTL I I NOSTTQVLFRLGMDLKS AKARRL LEFTKYEKKALAEFDLQDEDELOPADMTVKRY I KSLNLI SGR IHPHDA 573
SVB9-10_SQS1-C6Bhomolog1-1207 474 VVQVLYPFLCYFQYQWMEKEMAM LWAQLRQAVALSLSL SVKRS50GPAYSLTOTOTLFFVFTG I VFLFTL I I NOSTTQVLFRLGMDLKS AKARRL LEFTKYEKKALAEFDLQDEDELOPADMTVKRY I KSLNLI SGR IHPHDA 573

Cqinoa_SQS1-C6B1-1234 801 SDTSDNDF LDPMLKDMVRLLNVOVSA YWMLDEGR I TOSTANVLMSVDEALD VDHPEFLCDWKGLKNSVHF PKYRLLQGG I YPKKLVTF FVERLESACY I CAFLRAHRTARQGLHDF I DSE I ISAV I TESETEGEERKFL ELD 720
Cqinoa_SQS1-1211 578 SDTSDNDF LDPMLKDMVRLLNVOVSA YWMLDEGR I TOSTANVLMSVDEALD VDHPEFLCDWKGLKNSVHF PKYRLLQGG I YPKKLVTF FVERLESACY I CAFLRAHRTARQGLHDF I DSE I ISAV I TESETEGEERKFL ELD 727
Cberlandierutatae_SQS1-C6B1-1207 574 SDTSDNDF LDPMLKDMVRLLNVOVSA YWMLDEGR I TOSTANVLMSVDEALD VDHPEFLCDWKGLKNSVHF PKYRLLQGG I YPKKLVTF FVERLESACY I CAFLRAHRTARQGLHDF I DSE I ISAV I TESETEGEERKFL ELD 723
SVB9-10_SQS1-C6Bhomolog1-1207 574 SDTSDNDF LDPMLKDMVRLLNVOVSA YWMLDEGR I TOSTANVLMSVDEALD VDHPEFLCDWKGLKNSVHF PKYRLLQGG I YPKKLVTF FVERLESACY I CAFLRAHRTARQGLHDF I DSE I ISAV I TESETEGEERKFL ELD 723

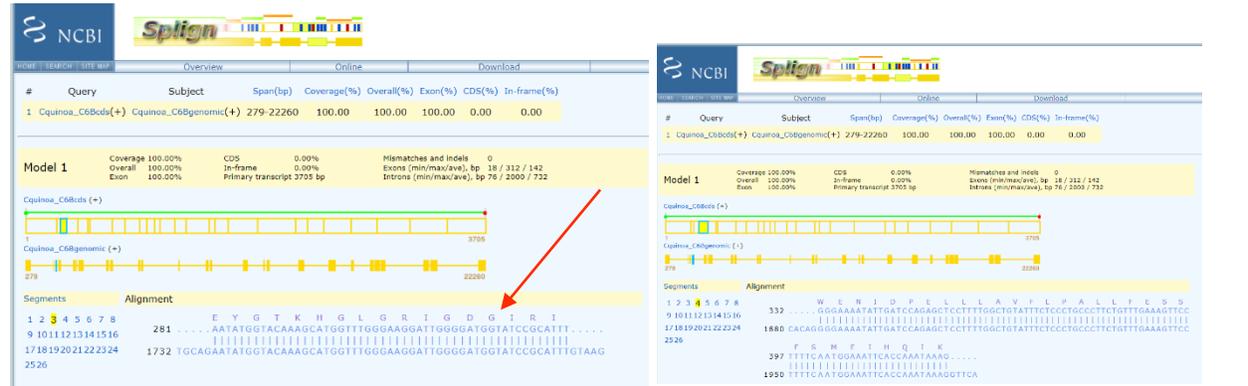
Cqinoa_SQS1-C6B1-1234 751 VRTTPEVLRVVKTRQVTVAVLQHLI EYIESLEKAG I LEKEMLHLDVAQDTDLKRLVRNPTVK I PKI OELI SMHFFLQALPSGVRDLV GSTKEEVKVRMTLYKEGQKPM I WLSI S6VWVWASVQRKKNHALHOTT FHTSTLGLYE 877
Cqinoa_SQS1-1211 728 VRTTPEVLRVVKTRQVTVAVLQHLI EYIESLEKAG I LEKEMLHLDVAQDTDLKRLVRNPTVK I PKI OELI SMHFFLQALPSGVRDLV GSTKEEVKVRMTLYKEGQKPM I WLSI S6VWVWASVQRKKNHALHOTT FHTSTLGLYE 873
Cberlandierutatae_SQS1-C6B1-1207 724 VRTTPEVLRVVKTRQVTVAVLQHLI EYIESLEKAG I LEKEMLHLDVAQDTDLKRLVRNPTVK I PKI OELI SMHFFLQALPSGVRDLV GSTKEEVKVRMTLYKEGQKPM I WLSI S6VWVWASVQRKKNHALHOTT FHTSTLGLYE 873
SVB9-10_SQS1-C6Bhomolog1-1207 724 VRTTPEVLRVVKTRQVTVAVLQHLI EYIESLEKAG I LEKEMLHLDVAQDTDLKRLVRNPTVK I PKI OELI SMHFFLQALPSGVRDLV GSTKEEVKVRMTLYKEGQKPM I WLSI S6VWVWASVQRKKNHALHOTT FHTSTLGLYE 873

Cqinoa_SQS1-C6B1-1234 801 VLIQKPYLDMITDSVAVCFYIETEKI LAALSDPAVEHFFWKESVILAVKLLPRV FEMMSQDMKLTAE RSTLNTLYRGTIEVPSHSI OF LLEGF I KSHLVEELI TSPAALWAGN5SFL QG6G5YKSTF LHDG5AVYTER 1050
Cqinoa_SQS1-1211 1028 VLIQKPYLDMITDSVAVCFYIETEKI LAALSDPAVEHFFWKESVILAVKLLPRV FEMMSQDMKLTAE RSTLNTLYRGTIEVPSHSI OF LLEGF I KSHLVEELI TSPAALWAGN5SFL QG6G5YKSTF LHDG5AVYTER 1027
Cberlandierutatae_SQS1-C6B1-1207 1024 VLIQKPYLDMITDSVAVCFYIETEKI LAALSDPAVEHFFWKESVILAVKLLPRV FEMMSQDMKLTAE RSTLNTLYRGTIEVPSHSI OF LLEGF I KSHLVEELI TSPAALWAGN5SFL QG6G5YKSTF LHDG5AVYTER 1023
SVB9-10_SQS1-C6Bhomolog1-1207 1024 VLIQKPYLDMITDSVAVCFYIETEKI LAALSDPAVEHFFWKESVILAVKLLPRV FEMMSQDMKLTAE RSTLNTLYRGTIEVPSHSI OF LLEGF I KSHLVEELI TSPAALWAGN5SFL QG6G5YKSTF LHDG5AVYTER 1023

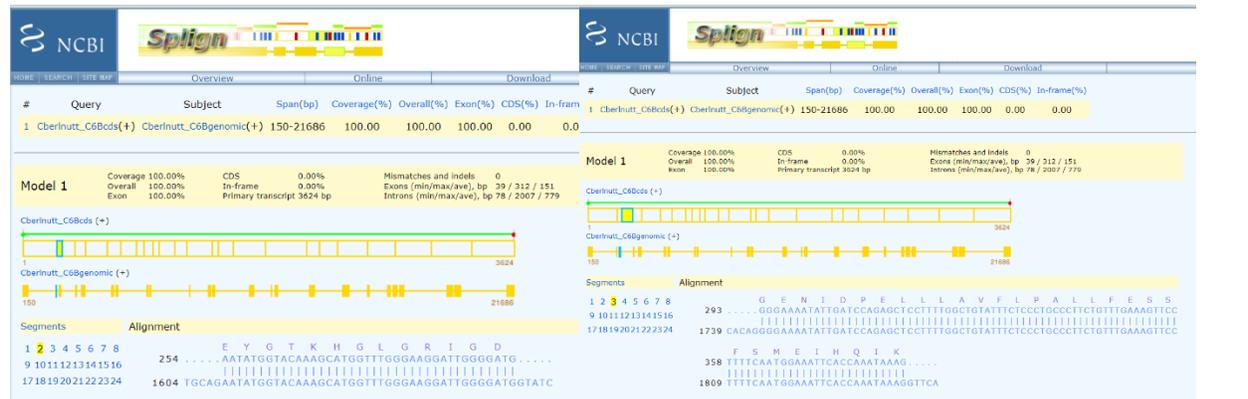
Cqinoa_SQS1-C6B1-1234 1051 ARVLLIDMVP I QADNTLLRRK5SLL LHDS5RSLNSRDHAGLLSPFENYKSHRPLDQGE I QDSONL SAKAMRL I YGSTARVPLRGLSFOGYSLGNPSHVRSYPOVP I QGOKRPLTSVKSESNTRKRLGEDVMREELL P QTHSRH 1100
Cqinoa_SQS1-1211 1128 ARVLLIDMVP I QADNTLLRRK5SLL LHDS5RSLNSRDHAGLLSPFENYKSHRPLDQGE I QDSONL SAKAMRL I YGSTARVPLRGLSFOGYSLGNPSHVRSYPOVP I QGOKRPLTSVKSESNTRKRLGEDVMREELL P QTHSRH 1177
Cberlandierutatae_SQS1-C6B1-1207 11024 ARVLLIDMVP I QADNTLLRRK5SLL LHDS5RSLNSRDHAGLLSPFENYKSHRPLDQGE I QDSONL SAKAMRL I YGSTARVPLRGLSFOGYSLGNPSHVRSYPOVP I QGOKRPLTSVKSESNTRKRLGEDVMREELL P QTHSRH 1173
SVB9-10_SQS1-C6Bhomolog1-1207 11024 ARVLLIDMVP I QADNTLLRRK5SLL LHDS5RSLNSRDHAGLLSPFENYKSHRPLDQGE I QDSONL SAKAMRL I YGSTARVPLRGLSFOGYSLGNPSHVRSYPOVP I QGOKRPLTSVKSESNTRKRLGEDVMREELL P QTHSRH 1173

Cqinoa_SQS1-C6B1-1234 1201 P5RVVDDSSSESGGEEVIVR I D5PSKLSFROAP 1234
Cqinoa_SQS1-1211 1178 P5RVVDDSSSESGGEEVIVR I D5PSKLSFROAP 1211
Cberlandierutatae_SQS1-C6B1-1207 1174 P5RVVDDSSSESGGEEVIVR I D5PSKLSFROAP 1207
SVB9-10_SQS1-C6Bhomolog1-1207 1174 P5RVVDDSSSESGGEEVIVR I D5PSKLSFROAP 1207
    
```

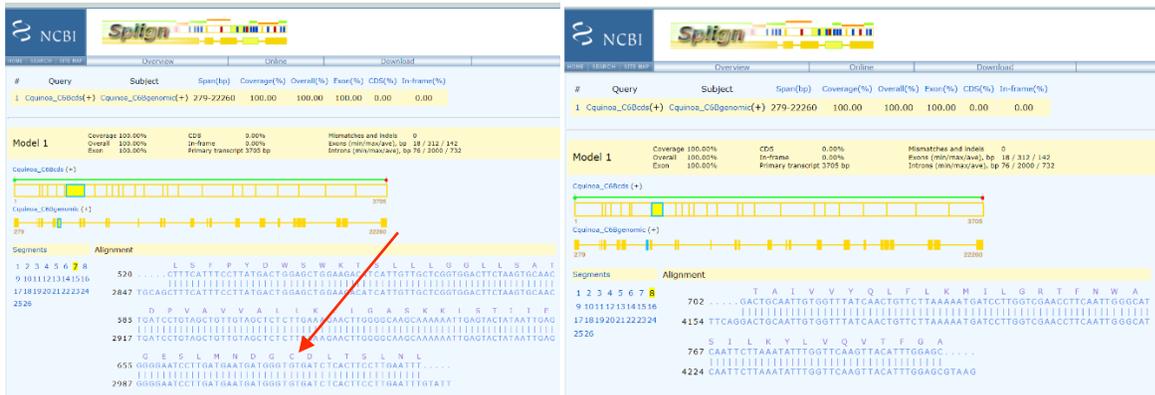
B.



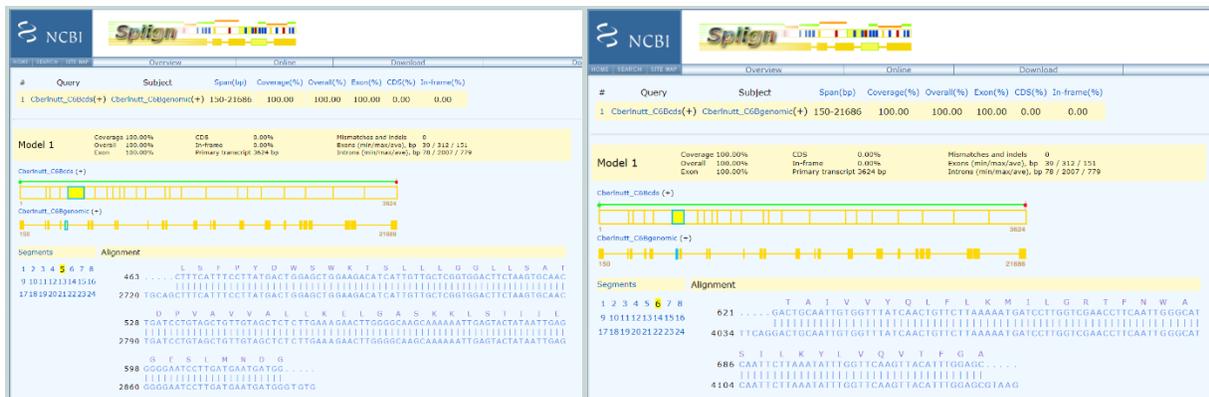
C.



D.



E.



F.

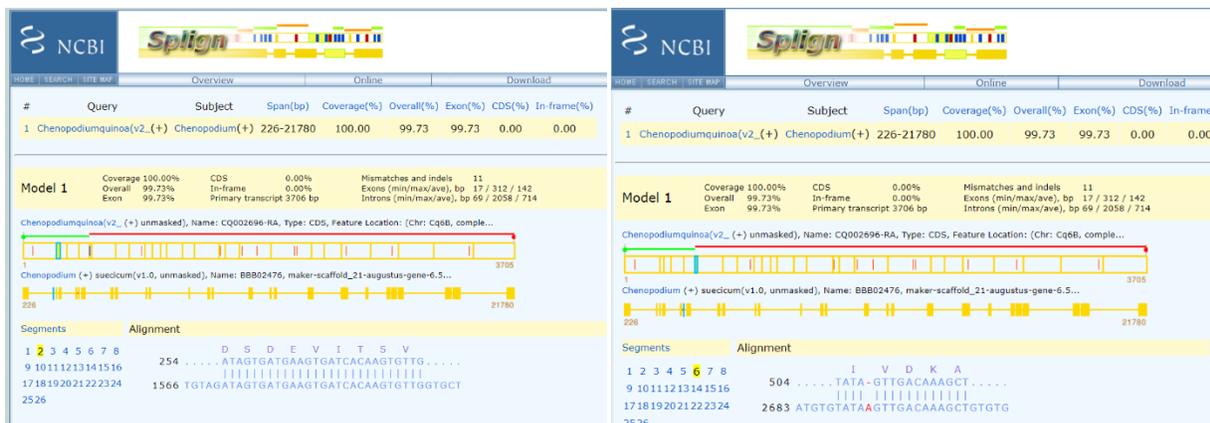


Figure S9. Gene models of *SOS1* sub-genome B. In all panels, the sequence whose name begins SV89-10 is from the *de novo* assembly reported here. A. Multiple sequence alignment of gene models from CoGe database for *C. quinoa* sub-genome B, *C. berlandieri nuttaliae*, and *C. suecicum* (the sub-genome B

related diploid), with the gene model assembled from our assembly scaffolds with greatest similarity to the quinoa sub-genome B homolog. B. *C. quinoa SOS1* C6B homolog, exon 3/4 junction, red arrow indicates GT splice donor site used in other species gene annotations. C. *C. berlandieri nuttaliae SOS1* C6B homolog exon 2/3 junction. D. *C. quinoa SOS1* C6B homolog, exon 7/8 junction, red arrow indicates GT splice donor site used in other species gene annotations. E. *C. berlandieri nuttaliae SOS1* C6B homolog, exon 6/7 junction. F. Possible alternative exons in *C. suecicum* genomic sequence, using the *C. quinoa SOS1* C6B homolog as reference.

Samuels Genomic sequence of Canadian *Chenopodium berlandieri*

A.

```

Qqinoa_SQS1_C6A1-1281 1 MAAYR IAVPFLMLAEVQAEITAPMNSTI SASAMAEESNPDAVIFFGVSLI I O IACRHLRQTRVPYTVALL I I O I O L O S L D E V I T S E Y O T K H O L R F O D O I R I W E N I D P E L L A V L F P A L L F E S S F S M E I H O I K R C A A O M I L L 150
QaHidicauLe_SQS1-1211 1 MAAYR IAVPFLMLAEVQAEITAPMNSTI SASAMAEESNPDAVIFFGVSLI I O IACRHLRQTRVPYTVALL I I O I O L O S L D E V I T S E Y O T K H O L R F O D O I R I W E N I D P E L L A V L F P A L L F E S S F S M E I H O I K R C A A O M I L L 141
CbeHmat_SQS1_C6A1-1152 1 MAAYR IAVPFLMLAEVQAEITAPMNSTI SASAMAEESNPDAVIFFGVSLI I O IACRHLRQTRVPYTVALL I I O I O L O S L D E V I T S E Y O T K H O L R F O D O I R I W E N I D P E L L A V L F P A L L F E S S F S M E I H O I K R C A A O M I L L 141
SVB9-1Q_SQS1_C6A1-1152 1 MAAYR IAVPFLMLAEVQAEITAPMNSTI SASAMAEESNPDAVIFFGVSLI I O IACRHLRQTRVPYTVALL I I O I O L O S L D E V I T S E Y O T K H O L R F O D O I R I W E N I D P E L L A V L F P A L L F E S S F S M E I H O I K R C A A O M I L L 141

Qqinoa_SQS1_C6A1-1281 151 A G P G V L I S T F L C L A A L K V K V D K L S F P Y D W S W K T S L L G O L L S A T D P V A V A L L K E L G A S K K L S T I I E G E S L M D O G D L T S L N L Y S M F L A E Y I D F V R T I V T A I V V Y Q L F K M I L O R T F N W A S I L K Y L V Q V S F G A V O F O I A F O I A S V L W L 300
QaHidicauLe_SQS1-1211 142 A G P G V L I S T F L C L A A L K V K V D K L S F P Y D W S W K T S L L G O L L S A T D P V A V A L L K E L G A S K K L S T I I E G E S L M D O G D L T S L N L Y S M F L A E Y I D F V R T I V T A I V V Y Q L F K M I L O R T F N W A S I L K Y L V Q V S F G A V O F O I A F O I A S V L W L 295
CbeHmat_SQS1_C6A1-1152 142 A G P G V L I S T F L C L A A L K V K V D K L S F P Y D W S W K T S L L G O L L S A T D P V A V A L L K E L G A S K K L S T I I E G E S L M D O G D L T S L N L Y S M F L A E Y I D F V R T I V T A I V V Y Q L F K M I L O R T F N W A S I L K Y L V Q V S F G A V O F O I A F O I A S V L W L 200
SVB9-1Q_SQS1_C6A1-1152 142 A G P G V L I S T F L C L A A L K V K V D K L S F P Y D W S W K T S L L G O L L S A T D P V A V A L L K E L G A S K K L S T I I E G E S L M D O G D L T S L N L Y S M F L A E Y I D F V R T I V T A I V V Y Q L F K M I L O R T F N W A S I L K Y L V Q V S F G A V O F O I A F O I A S V L W L 200

Qqinoa_SQS1_C6A1-1281 301 9 F I F N D T I I E I T L T L A V S V A A F T V I W K R D A R G S A G S E G K I R T K R R G I M R E N E E D R K I S S V A G E G A D V S Q L T V M T L O M F Y A A A R T A F K O E S O O S L H F W E M V A Y I A N T L I F I L S G A V I A G O V L S D N I F E N H N G A S V S K L S N S V 450
QaHidicauLe_SQS1-1211 280 9 F I F N D T I I E I T L T L A V S V A A F T V I W K R D A R G S A G S E G K I R T K R R G I M R E N E E D R K I S S V A G E G A D V S Q L T V M T L O M F Y A A A R T A F K O E S O O S L H F W E M V A Y I A N T L I F I L S G A V I A G O V L S D N I F E N H N G A S V S K L S N S V 380
CbeHmat_SQS1_C6A1-1152 201 9 F I F N D T I I E I T L T L A V S V A A F T V I W K R D A R G S A G S E G K I R T K R R G I M R E N E E D R K I S S V A G E G A D V S Q L T V M T L O M F Y A A A R T A F K O E S O O S L H F W E M V A Y I A N T L I F I L S G A V I A G O V L S D N I F E N H N G A S V S K L S N S V 357
SVB9-1Q_SQS1_C6A1-1152 201 9 F I F N D T I I E I T L T L A V S V A A F T V I W K R D A R G S A G S E G K I R T K R R G I M R E N E E D R K I S S V A G E G A D V S Q L T V M T L O M F Y A A A R T A F K O E S O O S L H F W E M V A Y I A N T L I F I L S G A V I A G O V L S D N I F E N H N G A S V S K L S N S V 357

Qqinoa_SQS1_C6A1-1281 451 S K Y L N O L S Y L M E A G H N R . . . E D N H S E M H U G T A W O Y L I L L Y V V L V A R G V V O V L Y P F L C Y F G Y M E N K E A M I L V W A G L R G A V A L S L S L S V K R S S D P A Y L S T Q T G T L F V F F T O I V F L T L I I N O S T T Q F L R F L O M D K L S K A K R R I 597
QaHidicauLe_SQS1-1211 461 0 0 0 I E K R I Q . . . L N T L O H L S D L S V G V F F E L D S D G T A W O Y L I L L Y V V L V A R G V V O V L Y P F L C Y F G Y M E N K E A M I L V W A G L R G A V A L S L S L S V K R S S D P A Y L S T Q T G T L F V F F T O I V F L T L I I N O S T T Q F L R F L O M D K L S K A K R R I 528
CbeHmat_SQS1_C6A1-1152 358 . . . . . 0 T A W O Y L I L L Y V V L V A R G V V O V L Y P F L C Y F G Y M E N K E A M I L V W A G L R G A V A L S L S L S V K R S S D P A Y L S T Q T G T L F V F F T O I V F L T L I I N O S T T Q F L R F L O M D K L S K A K R R I 474
SVB9-1Q_SQS1_C6A1-1152 358 . . . . . 0 T A W O Y L I L L Y V V L V A R G V V O V L Y P F L C Y F G Y M E N K E A M I L V W A G L R G A V A L S L S L S V K R S S D P A Y L S T Q T G T L F V F F T O I V F L T L I I N O S T T Q F L R F L O M D K L S K A K R R I 474

Qqinoa_SQS1_C6A1-1281 598 L E F T K Y E M E K K A L E A F O D L G E D E L O P A D W P T V K R Y I K S L N S I D G D R I H P H D A S D N O F L D P M N L K O M R V R L L N G O A A Y W A M L D E G R I T O S T A N V L M O S V D E A L D S V D H E P L C D W K G L K N S V H F P K Y R L L O G O I P K K L V T F T V E R L E 747
QaHidicauLe_SQS1-1211 529 L E F T K Y E M E K K A L E A F O D L G E D E L O P A D W P T V K R Y I K S L N S I D G D R I H P H D A S D N O F L D P M N L K O M R V R L L N G O A A Y W A M L D E G R I T O S T A N V L M O S V D E A L D S V D H E P L C D W K G L K N S V H F P K Y R L L O G O I P K K L V T F T V E R L E 678
CbeHmat_SQS1_C6A1-1152 475 L E F T K Y E M E K K A L E A F O D L G E D E L O P A D W P T V K R Y I K S L N S I D G D R I H P H D A S D N O F L D P M N L K O M R V R L L N G O A A Y W A M L D E G R I T O S T A N V L M O S V D E A L D S V D H E P L C D W K G L K N S V H F P K Y R L L O G O I P K K L V T F T V E R L E 624
SVB9-1Q_SQS1_C6A1-1152 475 L E F T K Y E M E K K A L E A F O D L G E D E L O P A D W P T V K R Y I K S L N S I D G D R I H P H D A S D N O F L D P M N L K O M R V R L L N G O A A Y W A M L D E G R I T O S T A N V L M O S V D E A L D S V D H E P L C D W K G L K N S V H F P K Y R L L O G O I P K K L V T F T V E R L E 624

Qqinoa_SQS1_C6A1-1281 748 S A C Y I C A A F L R A H R T A R G O L H D F I O Y S E I S S A V I T E S E T E G E A R K F L E D V R T T F P E L V R V K K T R O V T Y A V L Q H I E Y I E S L E K A G I L E E K E M L H L D A V Q T D L K R L V R N P P T V K I P K I G E L I S M H F L G A L P S O V R D L L V G S T K E E V K 897
QaHidicauLe_SQS1-1211 679 S A C Y I C A A F L R A H R T A R G O L H D F I O Y S E I S S V I T E S E T E G E A R N F L E D V R T T F P E L V R V K K T R O V T Y A V L Q H I E Y I E S L E K A G I L E E K E M L H L D A V Q T D L K R L V R N P P T V K I P K I G E L I S M H F L G A L P S O V R D L L V G S T K E E V K 828
CbeHmat_SQS1_C6A1-1152 625 S A C Y I C A A F L R A H R T A R G O L H D F I O Y S E I S S A V I T E S E T E G E A R K F L E D V R T T F P E . . . . . T R O V T Y A V L Q H I E Y I E S L E K A G I L E E K E M L H L D A V Q T D L K R L V R N P P T V K I P K I G E L I S M H F L G A L P S O V R D L L V G S T K E E V K 708
SVB9-1Q_SQS1_C6A1-1152 625 S A C Y I C A A F L R A H R T A R G O L H D F I O Y S E I S S A V I T E S E T E G E A R N F L E D V R T T F P E . . . . . T R O V T Y A V L Q H I E Y I E S L E K A G I L E E K E M L H L D A V Q T D L K R L V R N P P T V K I P K I G E L I S M H F L G A L P S O V R D L L V G S T K E E V K 695

Qqinoa_SQS1_C6A1-1281 898 R O M T L Y K E O K P N O I W L I S N O V V K A C K V K K K H A L H O T V T H S T L O L Y E V L I O K P Y L C D M I T O S V A V C F Y I E T E K I L A A L O S P A V E N F H F W E S V I L A K V L L P R V F E N S M O D M R K L T A E R S T L N T Y L R O E T I E V S H S I O F L L E O F V 1047
QaHidicauLe_SQS1-1211 829 R O M T L Y K E O K P N O I W L I S N O V V K A C K V K K K H A L H O T V T H S T L O L Y E V L I O K P Y L C D M I T O S V A V C F Y I E T E K I L A A L O S P A V E N F H F W E S V I L A K V L L P R V F E N S M O D M R K L T A E R S T L N T Y L R O E T I E V S H S I O F L L E O F V 977
CbeHmat_SQS1_C6A1-1152 769 R O M T L Y K E O K P N O I W L I S N O V V K A C K V K K K H A L H O T V T H S T L O L Y E V L I O K P Y L C D M I T O S V A V C F Y I E T E K I L A A L O S P A V E N F H F W E S V I L A K V L L P R V F E N S M O D M R K L T A E R S T L N T Y L R O E T I E V S H S I O F L L E O F V 918
SVB9-1Q_SQS1_C6A1-1152 769 R O M T L Y K E O K P N O I W L I S N O V V K A C K V K K K H A L H O T V T H S T L O L Y E V L I O K P Y L C D M I T O S V A V C F Y I E T E K I L A A L O S P A V E N F H F W E S V I L A K V L L P R V F E N S M O D M R K L T A E R S T L N T Y L R O E T I E V S H S I O F L L E O F V 918

Qqinoa_SQS1_C6A1-1281 1048 K S H P L A E E L I P S P A A L W P A G O N S F L S O E S O Y K S T S F L H O O T S Y V V E T R A R V L I D M P V I Q A D N T L L R R K S L L L H D O S S R L S S R D H A O L L S W P E N O F K S H G L P D O G E I D D S O N L S A K A M R L S I Y G S T A K O V P V R L S F O G Y T L O N R 1107
QaHidicauLe_SQS1-1211 1128 K S H P L A E E L I P S P A A L W P A G O N S F L S O E S O Y K S T S F L H O O T S Y V V E T R A R V L I D M P V I Q A D N T L L R R K S L L L H D O S S R L S S R D H A O L L S W P E N O F K S H G L P D O G E I D D S O N L S A K A M R L S I Y G S T A E D V P V R L S F O G Y T L O N R 1127
CbeHmat_SQS1_C6A1-1152 919 K S H P L A E E L I P S P A A L W P A G O N S F L S O E S O Y K S T S F L H O O T S Y V V E T R A R V L I D M P V I Q A D N T L L R R K S L L L H D O S S R L S S R D H A O L L S W P E N O F K S H G L P D O G E I D D S O N L S A K A M R L S I Y G S T A K O V P V R L S F O G Y T L O N R 1008
SVB9-1Q_SQS1_C6A1-1152 919 K S H P L A E E L I P S P A A L W P A G O N S F L S O E S O Y K S T S F L H O O T S Y V V E T R A R V L I D M P V I Q A D N T L L R R K S L L L H D O S S R L S S R D H A O L L S W P E N O F K S H G L P D O G E I D D S O N L S A K A M R L S I Y G S T A K O V P V R L S F O G Y T L O N R 1008

Qqinoa_SQS1_C6A1-1281 1108 S H V R S Y P Q V P I O G K Q L P L T S V K S E G S N T V R K R L G E D V M R E E L L P P T H S R M P S R A V D S S E S G E D E V F V R I D S P S K L S F R A P 1281
QaHidicauLe_SQS1-1211 1128 S H V R S Y P Q V P I O G K Q L P L T S V K S E G S N T V R K R L G E D V M R E E L L P P T H S R M P S R A V D S S E S G E D E V F V R I D S P S K L S F R A P 1211
CbeHmat_SQS1_C6A1-1152 1009 S H V R S Y P Q V P I O G K Q L P L T S V K S E G S N T V R K R L G E D V M R E E L L P P T H S R M P S R A V D S S E S G E D E V F V R I D S P S K L S F R A P 1152
SVB9-1Q_SQS1_C6A1-1152 1009 S H V R S Y P Q V P I O G K Q L P L T S V K S E G S N T V R K R L G E D V M R E E L L P P T H S R M P S R A V D S S E S G E D E V F V R I D S P S K L S F R A P 1152
    
```

B.

NCBI SnpSift

Overview Online Download

#	Query	Subject	Span(bp)	Coverage(%)	Overall(%)	Exon(%)	CDS(%)	In-frame(%)
1	Chenopodium(+)	scaffold8627rc_79067rc_33144rc(+)	12784-33751	100.00	99.19	99.19	0.00	0.00

Model 1

Coverage	CDS	Mismatches and indels
100.00%	0.00%	28
99.19%	0.00%	Exons (min/max/ave), bp 45 / 312 / 150
99.19%	Primary transcript 3459 bp	Introns (min/max/ave), bp 51 / 2112 / 796

Chenopodium (+) berlandieri subsp. nuttalliae(v2, unmasked), Name: Cb041956-RA, Type: CDS, F..

1 scaffold8627rc_79067rc_33144rc (+) 3459

12784 33751

Segments Alignment

Segment	Alignment
1	G Y S E I S S A V I T E S E T E G E E A R K
2	1946 G T T A C A G T G A G A T A T C T T C T G C T G C A T C A C T G A A A G T G A G A C A G A A G G A G A G G A A A G C A A G A A A A
3	1516 17 18 19 20 21
4	27283 T A C A G G T T A C A G T G A G A T A T C T T C T G C T G C A T C G C T G A A A G T G A G A C A G A A G G A T A G G A A G C A A G A A A A
5	2223
6	F L E D V R T T F P E
7	2011 T T T T T G G A A G A T G T C C G C A C T A C T T T C C T G A G
8	27353 T T T T T G G A A G A T G T C C G C A C T A C T T T C C T G A G G T T A T

C.

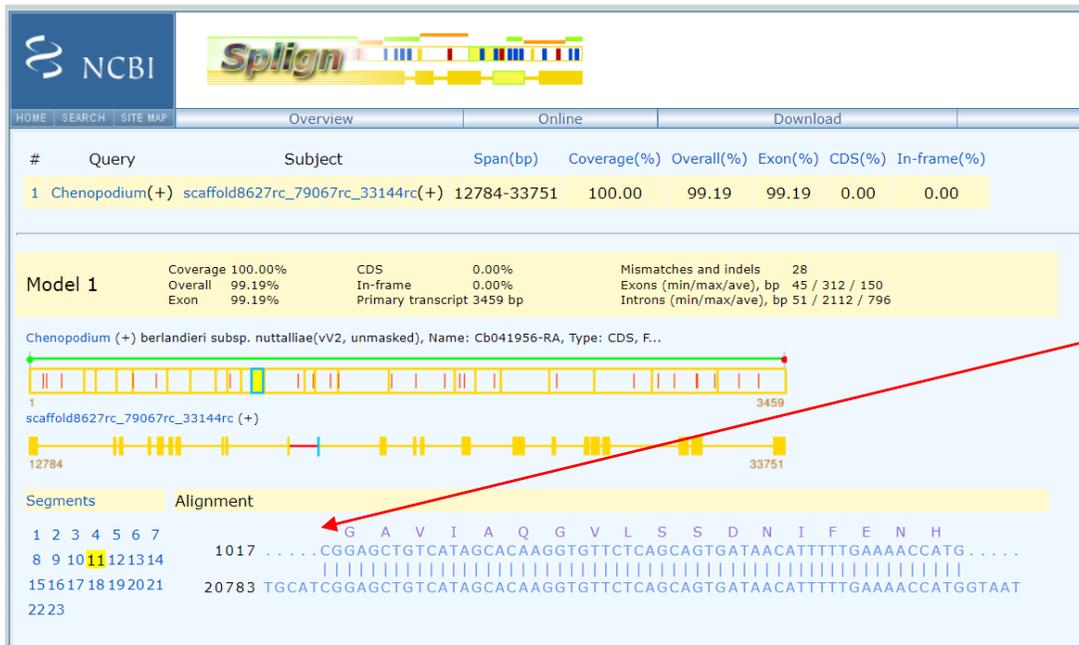


Figure S10. Gene models of *SOS1* sub-genome A. In all panels, the sequence whose name begins SV89-10 is from the *de novo* assembly reported here. A. Multiple sequence alignment of gene models from CoGe database for *C. quinoa* sub-genome A, *C. berlandieri nuttalliae*, and *C. pallidicaule* (the sub-genome A related diploid), with the gene model created from our assembly scaffolds with greatest similarity to the quinoa sub-genome A homolog. B. *C. berlandieri nuttalliae* *SOS1* homolog C6A coding sequence aligned to *C. berlandieri de novo* assembly, exon 17. In frame stop codon in *de novo* assembly indicated with red arrow. C. *C. berlandieri nuttalliae* *SOS1* homolog C6A coding sequence aligned to *C. berlandieri de novo* assembly, exon 11. Mutation in splice acceptor site in *de novo* assembly indicated with red arrow.

Samuels Genomic sequence of Canadian *Chenopodium berlandieri*

A.

```

Quinoa_CQ029119_C5A/1-362      1  . . . . . MDNYLVGNKIGKNIKRSPLHQPINFVNSGTNNNGNNNNNTARQOPQPVYNIKSKDFRSIVQQLTGSRSRQDPLRPPPNPKPHSMRLQIRPPLN  96
Galiliculae_CQ000195/1-372    1  MDNSMNSKRRPNDYLVGNKIGKNIKRSPLHQPINFVNSGTNNNGNNNNNTARQOPQPVYNIKSKDFRSIVQQLTGSRSRQDPLRPPPNPKPHSMRLQIRPPLN  100
Cberlandiennuttaliae_Cb039629_C5A/1-372  1  MDNSMNSKRRPNDYLVGNKIGKNIKRSPLHQPINFVNSGTNNNGNNNNNTARQOPQPVYNIKSKDFRSIVQQLTGSRSRQDPLRPPPNPKPHSMRLQIRPPLN  106
Scaffold_7576/1-372          1  MDNSMNSKRRPNDYLVGNKIGKNIKRSPLHQPINFVNSGTNNNGNNNNNTARQOPQPVYNIKSKDFRSIVQQLTGSRSRQDPLRPPPNPKPHSMRLQIRPPLN  102

Quinoa_CQ029119_C5A/1-362      97  INRRPMPVVAAPAFMQMPPFQGGHMAAPQVVFNRRPFPQGGSRPMQGFAPGDAWGNFALSPISAYMQQLQSCIIGDGGSRPQFHQOPMPPGQYQTOPM  206
Galiliculae_CQ000195/1-372    107  INRRPMPVVAAPAFMQMPPFQGGHMAAPQVVFNRRPFPQGGSRPMQGFAPGDAWGNFALSPISAYMQQLQSCIIGDGGSRPQFHQOPMPPGQYQTOPM  212
Cberlandiennuttaliae_Cb039629_C5A/1-372  107  INRRPMPVVAAPAFMQMPPFQGGHMAAPQVVFNRRPFPQGGSRPMQGFAPGDAWGNFALSPISAYMQQLQSCIIGDGGSRPQFHQOPMPPGQYQTOPM  212
Scaffold_7576/1-372          107  INRRPMPVVAAPAFMQMPPFQGGHMAAPQVVFNRRPFPQGGSRPMQGFAPGDAWGNFALSPISAYMQQLQSCIIGDGGSRPQFHQOPMPPGQYQTOPM  212

Quinoa_CQ029119_C5A/1-362      203  QDAPGHIQSPQMHHPGPHGPHGQAOVQHPMQLDVGHVQAOHPSPGLLNPNNMPPGSLPSPGTNGPALLPSTSTQGLLSPNGFFNMSPTRRSPHPLLSPGS  308
Galiliculae_CQ000195/1-372    213  QDAPGHIQSPQMHHPGPHGPHGQAOVQHPMQLDVGHVQAOHPSPGLLNPNNMPPGSLPSPGTNGPALLPSTSTQGLLSPNGFFNMSPTRRSPHPLLSPGS  318
Cberlandiennuttaliae_Cb039629_C5A/1-372  213  QDAPGHIQSPQMHHPGPHGPHGQAOVQHPMQLDVGHVQAOHPSPGLLNPNNMPPGSLPSPGTNGPALLPSTSTQGLLSPNGFFNMSPTRRSPHPLLSPGS  318
Scaffold_7576/1-372          213  QDAPGHIQSPQMHHPGPHGPHGQAOVQHPMQLDVGHVQAOHPSPGLLNPNNMPPGSLPSPGTNGPALLPSTSTQGLLSPNGFFNMSPTRRSPHPLLSPGS  318

Quinoa_CQ029119_C5A/1-362      300  FPPLLSNFAFSFGQSGILGPGPPPLSPGYAFPLSPGLFMSSPRWRDPSF  362
Galiliculae_CQ000195/1-372    310  FPPLLSNFAFSFGQSGILGPGPPPLSPGYAFPLSPGLFMSSPRWRDPSF  372
Cberlandiennuttaliae_Cb039629_C5A/1-372  310  FPPLLSNFAFSFGQSGILGPGPPPLSPGYAFPLSPGLFMSSPRWRDPSF  372
Scaffold_7576/1-372          310  FPPLLSNFAFSFGQSGILGPGPPPLSPGYAFPLSPGLFMSSPRWRDPSF  372
    
```

B.

```

Scaffold_1360/1-377          1  MDNSMNSKRRPNDYLVGNKIGKNIKRSPLHQPINFVNSGTNNNGNNNNNTARQOPQPVYNIKSKDFRSIVQQLTGSRSRQDPLRPPPNPKPHSMRLQIRPPLN  611
Galiliculae_CQ000195/1-372    140  MDNSMNSKRRPNDYLVGNKIGKNIKRSPLHQPINFVNSGTNNNGNNNNNTARQOPQPVYNIKSKDFRSIVQQLTGSRSRQDPLRPPPNPKPHSMRLQIRPPLN  618
Quinoa_CQ003751_C5B/1-374      140  MDNSMNSKRRPNDYLVGNKIGKNIKRSPLHQPINFVNSGTNNNGNNNNNTARQOPQPVYNIKSKDFRSIVQQLTGSRSRQDPLRPPPNPKPHSMRLQIRPPLN  618
Cberlandiennuttaliae_Cb007314_C5B/1-375  140  MDNSMNSKRRPNDYLVGNKIGKNIKRSPLHQPINFVNSGTNNNGNNNNNTARQOPQPVYNIKSKDFRSIVQQLTGSRSRQDPLRPPPNPKPHSMRLQIRPPLN  614

Scaffold_1360/1-377          152  PFGQGRSRPMQGFAPGDAWGNFALSPISAYMQQLQSCIIGDGGSRPQFHQOPMPPGQYQTOPMOPAPRHIGSQDPAFGQGGPHGQAGHPMHQVQGVQADQPPSGGLLNPNNMPPGSLPSPGSGNALLPSTSTQGLLSPN  302
Galiliculae_CQ000195/1-372    140  PFGQGRSRPMQGFAPGDAWGNFALSPISAYMQQLQSCIIGDGGSRPQFHQOPMPPGQYQTOPMOPAPRHIGSQDPAFGQGGPHGQAGHPMHQVQGVQADQPPSGGLLNPNNMPPGSLPSPGSGNALLPSTSTQGLLSPN  308
Quinoa_CQ003751_C5B/1-374      140  PFGQGRSRPMQGFAPGDAWGNFALSPISAYMQQLQSCIIGDGGSRPQFHQOPMPPGQYQTOPMOPAPRHIGSQDPAFGQGGPHGQAGHPMHQVQGVQADQPPSGGLLNPNNMPPGSLPSPGSGNALLPSTSTQGLLSPN  308
Cberlandiennuttaliae_Cb007314_C5B/1-375  150  PFGQGRSRPMQGFAPGDAWGNFALSPISAYMQQLQSCIIGDGGSRPQFHQOPMPPGQYQTOPMOPAPRHIGSQDPAFGQGGPHGQAGHPMHQVQGVQADQPPSGGLLNPNNMPPGSLPSPGSGNALLPSTSTQGLLSPN  300

Scaffold_1360/1-377          303  FFFNMSPTRRSPHPLLSPGSFPPLSPNFAFSFGQSGILGPGPPPLSPGYAFPLSPGLFMSSPRWRDPSF  377
Galiliculae_CQ000195/1-372    300  FFFNMSPTRRSPHPLLSPGSFPPLSPNFAFSFGQSGILGPGPPPLSPGYAFPLSPGLFMSSPRWRDPSF  374
Quinoa_CQ003751_C5B/1-374      300  FFFNMSPTRRSPHPLLSPGSFPPLSPNFAFSFGQSGILGPGPPPLSPGYAFPLSPGLFMSSPRWRDPSF  374
Cberlandiennuttaliae_Cb007314_C5B/1-375  301  FFFNMSPTRRSPHPLLSPGSFPPLSPNFAFSFGQSGILGPGPPPLSPGYAFPLSPGLFMSSPRWRDPSF  376
    
```

Figure S11. *IKU1* multiple sequence alignment. A. Homologs of quinoa *IKU1* sub-genome A. B. Homologs of quinoa *IKU1* sub-genome B. In both panels, the sequences whose names begin with ‘Scaffold’ are those of the *de novo* assembly reported here.

```

IKU2_subgenomeA/1-924      1  MRLEPPFTPLFLFLVLCFSNSASDELDLTLKSSLSQSNPKFLSTWIPSPNPTQVTSOICHTARN  VTIIELSKORLSOTLPPNAICSLPSLQKLAFOFNSLFTISDRIHCTELRYLDLQNNQFSQSPFNISLHRLEFLVYLNLS  100
Quinoa_CQ023322_C18/1-924    1  MPPQPPFTPLFLFLVLCFSNSASDELDLTLKSSLSQSNPKFLSTWIPSPNPTQVTSOICHTARN  VTIIELSKORLSOTLPPNAICSLPSLQKLAFOFNSLFTISDRIHCTELRYLDLQNNQFSQSPFNISLHRLEFLVYLNLS  100
Galiliculae_Cb01174/ntendedSpiraea_pallidifoliae/1-924  1  MPPQPPFTPLFLFLVLCFSNSASDELDLTLKSSLSQSNPKFLSTWIPSPNPTQVTSOICHTARN  VTIIELSKORLSOTLPPNAICSLPSLQKLAFOFNSLFTISDRIHCTELRYLDLQNNQFSQSPFNISLHRLEFLVYLNLS  100
Cberlandiennuttaliae_Cb01174/ntendedSpiraea_pallidifoliae/1-924  1  MPPQPPFTPLFLFLVLCFSNSASDELDLTLKSSLSQSNPKFLSTWIPSPNPTQVTSOICHTARN  VTIIELSKORLSOTLPPNAICSLPSLQKLAFOFNSLFTISDRIHCTELRYLDLQNNQFSQSPFNISLHRLEFLVYLNLS  100
Galiliculae_CQ013578/1-924    1  MPPQPPFTPLFLFLVLCFSNSASDELDLTLKSSLSQSNPKFLSTWIPSPNPTQVTSOICHTARN  VTIIELSKORLSOTLPPNAICSLPSLQKLAFOFNSLFTISDRIHCTELRYLDLQNNQFSQSPFNISLHRLEFLVYLNLS  100
Quinoa_CQ049324_C1A/1-924    1  MPPQPPFTPLFLFLVLCFSNSASDELDLTLKSSLSQSNPKFLSTWIPSPNPTQVTSOICHTARN  VTIIELSKORLSOTLPPNAICSLPSLQKLAFOFNSLFTISDRIHCTELRYLDLQNNQFSQSPFNISLHRLEFLVYLNLS  100
Cberlandiennuttaliae_Cb01174/ntendedSpiraea_pallidifoliae/1-924  1  MPPQPPFTPLFLFLVLCFSNSASDELDLTLKSSLSQSNPKFLSTWIPSPNPTQVTSOICHTARN  VTIIELSKORLSOTLPPNAICSLPSLQKLAFOFNSLFTISDRIHCTELRYLDLQNNQFSQSPFNISLHRLEFLVYLNLS  100

IKU2_subgenomeA/1-924      151  FFSKGFPNWLSNMTGLIHLNSLQNFDEFTFPQVTEKLNKLLVNTSISQSIIPPEIIONSELSNLSNLSNLSIPIEISKLLKWLQEMVNNVLTQVIVPQFNLTNLRKFDASNIGDLSSEKLYNLWLIQVYQNEF  301
Galiliculae_Cb01174/ntendedSpiraea_pallidifoliae/1-924  151  FFSKGFPNWLSNMTGLIHLNSLQNFDEFTFPQVTEKLNKLLVNTSISQSIIPPEIIONSELSNLSNLSNLSIPIEISKLLKWLQEMVNNVLTQVIVPQFNLTNLRKFDASNIGDLSSEKLYNLWLIQVYQNEF  301
Cberlandiennuttaliae_Cb01174/ntendedSpiraea_pallidifoliae/1-924  151  FFSKGFPNWLSNMTGLIHLNSLQNFDEFTFPQVTEKLNKLLVNTSISQSIIPPEIIONSELSNLSNLSNLSIPIEISKLLKWLQEMVNNVLTQVIVPQFNLTNLRKFDASNIGDLSSEKLYNLWLIQVYQNEF  301
Galiliculae_CQ013578/1-924    151  FFSKGFPNWLSNMTGLIHLNSLQNFDEFTFPQVTEKLNKLLVNTSISQSIIPPEIIONSELSNLSNLSNLSIPIEISKLLKWLQEMVNNVLTQVIVPQFNLTNLRKFDASNIGDLSSEKLYNLWLIQVYQNEF  301
Quinoa_CQ049324_C1A/1-924    151  FFSKGFPNWLSNMTGLIHLNSLQNFDEFTFPQVTEKLNKLLVNTSISQSIIPPEIIONSELSNLSNLSNLSIPIEISKLLKWLQEMVNNVLTQVIVPQFNLTNLRKFDASNIGDLSSEKLYNLWLIQVYQNEF  301
Cberlandiennuttaliae_Cb01174/ntendedSpiraea_pallidifoliae/1-924  151  FFSKGFPNWLSNMTGLIHLNSLQNFDEFTFPQVTEKLNKLLVNTSISQSIIPPEIIONSELSNLSNLSNLSIPIEISKLLKWLQEMVNNVLTQVIVPQFNLTNLRKFDASNIGDLSSEKLYNLWLIQVYQNEF  301

IKU2_subgenomeA/1-924      302  PAFEDFQKLVNLSVSNLTGOLPKLQSWTEFDIVDSTNSLSPFPDMQKKTATKLLLENKFSQIIPAYANTCTLFRVSDNLTGQVPAIWLQPNVNIIDIDADNKRQSIISDIAKAENLVOIYAANNELSQIPAEIA  402
Galiliculae_Cb01174/ntendedSpiraea_pallidifoliae/1-924  302  PAFEDFQKLVNLSVSNLTGOLPKLQSWTEFDIVDSTNSLSPFPDMQKKTATKLLLENKFSQIIPAYANTCTLFRVSDNLTGQVPAIWLQPNVNIIDIDADNKRQSIISDIAKAENLVOIYAANNELSQIPAEIA  402
Cberlandiennuttaliae_Cb01174/ntendedSpiraea_pallidifoliae/1-924  302  PAFEDFQKLVNLSVSNLTGOLPKLQSWTEFDIVDSTNSLSPFPDMQKKTATKLLLENKFSQIIPAYANTCTLFRVSDNLTGQVPAIWLQPNVNIIDIDADNKRQSIISDIAKAENLVOIYAANNELSQIPAEIA  402
Galiliculae_CQ013578/1-924    302  PAFEDFQKLVNLSVSNLTGOLPKLQSWTEFDIVDSTNSLSPFPDMQKKTATKLLLENKFSQIIPAYANTCTLFRVSDNLTGQVPAIWLQPNVNIIDIDADNKRQSIISDIAKAENLVOIYAANNELSQIPAEIA  402
Quinoa_CQ049324_C1A/1-924    302  PAFEDFQKLVNLSVSNLTGOLPKLQSWTEFDIVDSTNSLSPFPDMQKKTATKLLLENKFSQIIPAYANTCTLFRVSDNLTGQVPAIWLQPNVNIIDIDADNKRQSIISDIAKAENLVOIYAANNELSQIPAEIA  402
Cberlandiennuttaliae_Cb01174/ntendedSpiraea_pallidifoliae/1-924  302  PAFEDFQKLVNLSVSNLTGOLPKLQSWTEFDIVDSTNSLSPFPDMQKKTATKLLLENKFSQIIPAYANTCTLFRVSDNLTGQVPAIWLQPNVNIIDIDADNKRQSIISDIAKAENLVOIYAANNELSQIPAEIA  402

IKU2_subgenomeA/1-924      403  SLLLLDLNKKLQVPEISIKLKMNLNLYLQNNQFSQSPFNISLHRLEFLVYLNLSIPIEISKLLKWLQEMVNNVLTQVIVPQFNLTNLRKFDASNIGDLSSEKLYNLWLIQVYQNEF  603
Galiliculae_Cb01174/ntendedSpiraea_pallidifoliae/1-924  403  SLLLLDLNKKLQVPEISIKLKMNLNLYLQNNQFSQSPFNISLHRLEFLVYLNLSIPIEISKLLKWLQEMVNNVLTQVIVPQFNLTNLRKFDASNIGDLSSEKLYNLWLIQVYQNEF  603
Cberlandiennuttaliae_Cb01174/ntendedSpiraea_pallidifoliae/1-924  403  SLLLLDLNKKLQVPEISIKLKMNLNLYLQNNQFSQSPFNISLHRLEFLVYLNLSIPIEISKLLKWLQEMVNNVLTQVIVPQFNLTNLRKFDASNIGDLSSEKLYNLWLIQVYQNEF  603
Galiliculae_CQ013578/1-924    403  SLLLLDLNKKLQVPEISIKLKMNLNLYLQNNQFSQSPFNISLHRLEFLVYLNLSIPIEISKLLKWLQEMVNNVLTQVIVPQFNLTNLRKFDASNIGDLSSEKLYNLWLIQVYQNEF  603
Quinoa_CQ049324_C1A/1-924    403  SLLLLDLNKKLQVPEISIKLKMNLNLYLQNNQFSQSPFNISLHRLEFLVYLNLSIPIEISKLLKWLQEMVNNVLTQVIVPQFNLTNLRKFDASNIGDLSSEKLYNLWLIQVYQNEF  603
Cberlandiennuttaliae_Cb01174/ntendedSpiraea_pallidifoliae/1-924  403  SLLLLDLNKKLQVPEISIKLKMNLNLYLQNNQFSQSPFNISLHRLEFLVYLNLSIPIEISKLLKWLQEMVNNVLTQVIVPQFNLTNLRKFDASNIGDLSSEKLYNLWLIQVYQNEF  603

IKU2_subgenomeA/1-924      604  FVILVGLQALVALSQCIVLYTKLQVQKDEKRSFLKDDYVWMMKSYHLLTTEELDLSIKPEKLIQKGGQSVYKVLQKQKELAVKHWNDFSDQKLRPSTPMLDKKYSGRKKEFDMEKTLSSIRHVVVLYSITSDS  754
Galiliculae_Cb01174/ntendedSpiraea_pallidifoliae/1-924  604  FVILVGLQALVALSQCIVLYTKLQVQKDEKRSFLKDDYVWMMKSYHLLTTEELDLSIKPEKLIQKGGQSVYKVLQKQKELAVKHWNDFSDQKLRPSTPMLDKKYSGRKKEFDMEKTLSSIRHVVVLYSITSDS  754
Cberlandiennuttaliae_Cb01174/ntendedSpiraea_pallidifoliae/1-924  604  FVILVGLQALVALSQCIVLYTKLQVQKDEKRSFLKDDYVWMMKSYHLLTTEELDLSIKPEKLIQKGGQSVYKVLQKQKELAVKHWNDFSDQKLRPSTPMLDKKYSGRKKEFDMEKTLSSIRHVVVLYSITSDS  754
Galiliculae_CQ013578/1-924    604  FVILVGLQALVALSQCIVLYTKLQVQKDEKRSFLKDDYVWMMKSYHLLTTEELDLSIKPEKLIQKGGQSVYKVLQKQKELAVKHWNDFSDQKLRPSTPMLDKKYSGRKKEFDMEKTLSSIRHVVVLYSITSDS  754
Quinoa_CQ049324_C1A/1-924    604  FVILVGLQALVALSQCIVLYTKLQVQKDEKRSFLKDDYVWMMKSYHLLTTEELDLSIKPEKLIQKGGQSVYKVLQKQKELAVKHWNDFSDQKLRPSTPMLDKKYSGRKKEFDMEKTLSSIRHVVVLYSITSDS  754
Cberlandiennuttaliae_Cb01174/ntendedSpiraea_pallidifoliae/1-924  604  FVILVGLQALVALSQCIVLYTKLQVQKDEKRSFLKDDYVWMMKSYHLLTTEELDLSIKPEKLIQKGGQSVYKVLQKQKELAVKHWNDFSDQKLRPSTPMLDKKYSGRKKEFDMEKTLSSIRHVVVLYSITSDS  754

IKU2_subgenomeA/1-924      755  SLLVVEYMPNGSLWDRHSHKMLKALDWSRYEIALGAAGKLEYLHHQVYKPIHRDVKSSNILLDEFKLRPIADFGAKIPQVNTNDSSTHVIAGTHYIAPDSSIAPLKEADIKILKIALCTARPELRPTMRSVQMLDEAPDGL  905
Galiliculae_Cb01174/ntendedSpiraea_pallidifoliae/1-924  755  SLLVVEYMPNGSLWDRHSHKMLKALDWSRYEIALGAAGKLEYLHHQVYKPIHRDVKSSNILLDEFKLRPIADFGAKIPQVNTNDSSTHVIAGTHYIAPDSSIAPLKEADIKILKIALCTARPELRPTMRSVQMLDEAPDGL  905
Cberlandiennuttaliae_Cb01174/ntendedSpiraea_pallidifoliae/1-924  755  SLLVVEYMPNGSLWDRHSHKMLKALDWSRYEIALGAAGKLEYLHHQVYKPIHRDVKSSNILLDEFKLRPIADFGAKIPQVNTNDSSTHVIAGTHYIAPDSSIAPLKEADIKILKIALCTARPELRPTMRSVQMLDEAPDGL  905
Galiliculae_CQ013578/1-924    755  SLLVVEYMPNGSLWDRHSHKMLKALDWSRYEIALGAAGKLEYLHHQVYKPIHRDVKSSNILLDEFKLRPIADFGAKIPQVNTNDSSTHVIAGTHYIAPDSSIAPLKEADIKILKIALCTARPELRPTMRSVQMLDEAPDGL  905
Quinoa_CQ049324_C1A/1-924    755  SLLVVEYMPNGSLWDRHSHKMLKALDWSRYEIALGAAGKLEYLHHQVYKPIHRDVKSSNILLDEFKLRPIADFGAKIPQVNTNDSSTHVIAGTHYIAPDSSIAPLKEADIKILKIALCTARPELRPTMRSVQMLDEAPDGL  905
Cberlandiennuttaliae_Cb01174/ntendedSpiraea_pallidifoliae/1-924  755  SLLVVEYMPNGSLWDRHSHKMLKALDWSRYEIALGAAGKLEYLHHQVYKPIHRDVKSSNILLDEFKLRPIADFGAKIPQVNTNDSSTHVIAGTHYIAPDSSIAPLKEADIKILKIALCTARPELRPTMRSVQMLDEAPDGL  905

IKU2_subgenomeA/1-924      906  SIVGKDAKKIEESKEGF  924
Galiliculae_Cb01174/ntendedSpiraea_pallidifoliae/1-924  906  SIVGKDAKKIEESKEGF  924
Cberlandiennuttaliae_Cb01174/ntendedSpiraea_pallidifoliae/1-924  906  SIVGKDAKKIEESKEGF  924
Galiliculae_CQ013578/1-924    906  SIVGKDAKKIEESKEGF  925
Quinoa_CQ049324_C1A/1-924    906  SIVGKDAKKIEESKEGF  924
Cberlandiennuttaliae_Cb01174/ntendedSpiraea_pallidifoliae/1-924  906  SIVGKDAKKIEESKEGF  924
    
```

Figure S12. *IKU2* gene models. As described in the text, splice sites and initiating methionines were used for maximum consistency in all the models, although in several cases the genes are differently annotated in the CoGe database. The sequence whose name begins *IKU2_sub-genomeA* is that of the assembly reported here.

E.

NCBI Splign

HOME SEARCH SITE MAP Overview Online Download

#	Query	Subject	Span(bp)	Coverage(%)	Overall(%)	Exon(%)	CDS(%)	In-frame(%)
1	CqinoaCQ026523cds(+)	CqinoaCQ026523genomic(+)	5603-8967	100.00	100.00	100.00	0.00	0.00

Model 1

Coverage	100.00%	CDS	0.00%	Mismatches and indels	0
Overall	100.00%	In-frame	0.00%	Exons (min/max/ave), bp	19 / 683 / 233
Exon	100.00%	Primary transcript	1398 bp	Introns (min/max/ave), bp	77 / 826 / 394

CqinoaCQ026523cds (+) (v2, unmasked), Name: CQ026523-RA, Type: CDS, Feature Location: (Chr: Cq6A, c...)

CqinoaCQ026523genomic (+) (v2, unmasked), Name: CQ026523, snap_masked-Cq6A-processed-gene-194.46, Type:...

Segments Alignment

```

1 2 3 4 5 6
M E V K E I D R V V I T K P V A S R P S C T G
1 ATGGAGGTAAAAGAGATAGACCGAGTAGTCATAACAAAGCCAGTAGCTTCAAGGCCATCTTGTACTGGCT
5603 ATGGAGGTAAAAGAGATAGACCGAGTAGTCATAACAAAGCCAGTAGCTTCAAGGCCATCTTGTACTGGCT

F R S F S K L H S G V V N A S P P R T N A I S E
71 TTAGATCTTTCTCTAAACTTCATTCTGGTGTGTCAATGCATCACCACCACGTACGAATGCTATCTCTGA
5673 TTAGATCTTTCTCTAAACTTCATTCTGGTGTGTCAATGCATCACCACCACGTACGAATGCTATCTCTGA

M A V A A I K P K T V R V K S T I N Q S P H A
141 AATGGCAGTTGCTGCTATTAACCCAAAACAGTGAGGGTAAAGTCCACAATCAATCAGTCTCCACATGCC
5743 AATGGCAGTTGCTGCTATTAACCCAAAACAGTGAGGGTAAAGTCCACAATCAATCAGTCTCCACATGCC

K E I Q
211 AAAGAAATTCAG.....
5813 AAAGAAATTCAGGCAAG
    
```

F.

NCBI Splign

HOME SEARCH SITE MAP Overview Online Download

#	Query	Subject	Span(bp)	Coverage(%)	Overall(%)	Exon(%)	CDS(%)	In-frame(%)
1	Cqinoa_C8B_CQ001086_mRNA(+)	Cqinoa_C6A_CQ026523_genomic(+)	5597-8765	100.00	93.79	93.79	0.00	0.00

Model 1

Coverage	100.00%	CDS	0.00%	Mismatches and indels	87
Overall	93.79%	In-frame	0.00%	Exons (min/max/ave), bp	25 / 722 / 205
Exon	93.79%	Primary transcript	1323 bp	Introns (min/max/ave), bp	75 / 787 / 412

Cqinoa_C8B_CQ001086_mRNA (+)

Cqinoa_C6A_CQ026523_genomic (+)

Segments Alignment

```

1 2 3 4 5
M E V K D I D R V V I K P V A S R P S C
1 TTGGGAATGGAGGTAAAAGATATAGATCGAGTGTCTATAAAGCCAGTAGCTTCAAGGCCATCTTGTGA
5597 TTGGGAATGGAGGTAAAAGATATAGATCGAGTGTCTATAAAGCCAGTAGCTTCAAGGCCATCTTGTGA

T G F R S F S K L H S G A I N A S P P P A N A S
71 TTGGCTTGTAGATCTTTCTAAACTTCATTCTGGTGTGTCAATGCATCACCACCACGTACGAATGCTATCTCTGA
5667 TTGGCTTGTAGATCTTTCTAAACTTCATTCTGGTGTGTCAATGCATCACCACCACGTACGAATGCTATCTCTGA

S E M A V A A I K P K T V R V K S T I N Q S P
141 CTCTGGAAATGGAGGTAAAAGATATAGATCGAGTGTCTATAAAGCCAGTAGCTTCAAGGCCATCTTGTGA
5737 CTCTGGAAATGGAGGTAAAAGATATAGATCGAGTGTCTATAAAGCCAGTAGCTTCAAGGCCATCTTGTGA

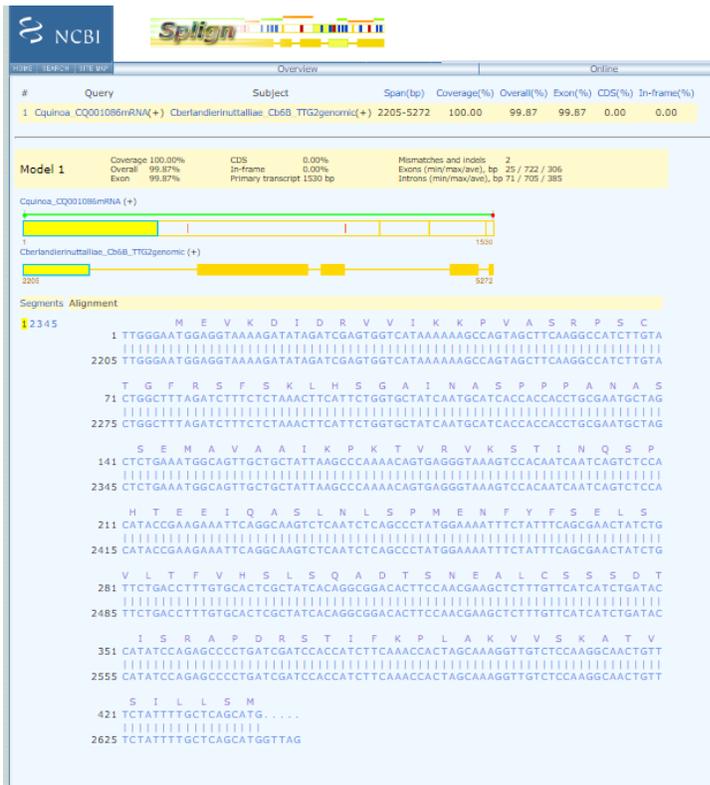
H P P S I Q A S L N L S P H S E R P F Y F S S A
211 CATACCGAAGAAATTCAGCCAGTCTCAATCTCACCCCTATGGA--AAATTTCTATTTAGCGAATCTATC
5807 CATACCGAAGAAATTCAGCCAGTCTCAATCTCACCCCTATGGA--AAATTTCTATTTAGCGAATCTATC

V L T F V H D L S Q A D T S N E A L C S S S D
279 TGTTCGACCTTTGTGCATCTGCTACAGGCGGACACTCCACAGGCTCTTGTTCATCTGAT
5874 ---CTGACCTTTGTGCATCTGCTACAGGCGGACACTCCACAGGCTCTTGTTCATCTGAT

T I S R A P D R E S T E F K P L A K V V S K A T
349 ACCATATCAGAGCCCTGATCGATCCACTCTTCAAACTACGAAAGGTTGCTCAGGCAACTG
5940 AACATATCAGAGCCCTGATCGATCCACTCTTCAAACTACGAAAGGTTGCTCAGGCAACTG

V S I L I D H
419 TTTCATTTTCTCAGCACTG....
6010 TTTCATTTTCTCAGCACTG....
    
```

G.



H.

Gpallidicaule_CP003660/1-465	1	MEVKEIDRVVITKPVASRPSCTOFRSF SKLHSGV VNASPPANA I SEMAVAA IKPKTVRVKSTINQSPHAKIQAN.....TGRAPCSSSDNISRVPDRSTVFK	100
CsueciumBBB189871-499	1	MEVKEIDRVVITKPVASRPSCTOFRSF SKLHSGA INASPPANASSEMVA AAIKPKTVRVKSTINQSPHTEEIQASLNLSPMNFYFSELSVLT FVHSLSQADTSNEALCSSDTSIRAPDRSTIFK	126
Cberlandier_ncafold273371-507	1	MEVKEIDRVVITKPVASRPSCTOFRSF SKLHSGA INASPPANASSEMVA AAIKPKTVRVKSTINQSPHTEEIQASLNLSPMNFYFSELSVLT FVHSLSQADTSNEALCSSDTSIRAPDRSTIFK	127
Quinoa_CQ001086_C68/1-507	1	MEVKEIDRVVITKPVASRPSCTOFRSF SKLHSGA INASPPANASSEMVA AAIKPKTVRVKSTINQSPHTEEIQASLNLSPMNFYFSELSVLT FVHSLSQADTSNEALCSSDTSIRAPDRSTIFK	127
Cberlandierinuttalliae_Cb002931_C68/1-507	1	MEVKEIDRVVITKPVASRPSCTOFRSF SKLHSGA INASPPANASSEMVA AAIKPKTVRVKSTINQSPHTEEIQASLNLSPMNFYFSELSVLT FVHSLSQADTSNEALCSSDTSIRAPDRSTIFK	127
Gpallidicaule_CP003660/1-465	101	PLAKVVS KATVSI LL SMV.....DVSARQKHLHFSOHSNLRVVELSES AKSPSMAQQMVEGEPDIMSSSIAGADRPSDDGYNWRKYGQKVKVGSYPRSYKCTHLKCPVKKKIER	214
CsueciumBBB189871-499	127	PLAKVVS KATVSI LL SMONSSVTHROALSOVNVPAQEKHLHFSOHSNLRVVELAESAKSPTMAQQMVEGESRIMASSITGADDPSCDDGYNWRKYGQKVKVGSYPRSYKCTHLKCPVKKKVER	253
Cberlandier_ncafold273371-507	126	PLAKVVS KATVSI LL SMONSSVTHROALSOVNVPAQEKHLHFSOHSNLRVVELAESAKSPTMAQQMVEGESRIMASSITGADDPSCDDGYNWRKYGQKVKVGSYPRSYKCTHLKCPVKKKVER	254
Quinoa_CQ001086_C68/1-507	128	PLAKVVS KATVSI LL SMONSSVTHROALSOVNVPAQEKHLHFSOHSNLRVVELAESAKSPTMAQQMVEGESRIMASSITGADDPSCDDGYNWRKYGQKVKVGSYPRSYKCTHLKCPVKKKVER	254
Cberlandierinuttalliae_Cb002931_C68/1-507	128	PLAKVVS KATVSI LL SMONSSVTHROALSOVNVPAQEKHLHFSOHSNLRVVELAESAKSPTMAQQMVEGESRIMASSITGADDPSCDDGYNWRKYGQKVKVGSYPRSYKCTHLKCPVKKKVER	254
Gpallidicaule_CP003660/1-465	215	SLDQGIT EIVYKGEHNNPKPDHAKRSSGVHGOELFAQTNQDPNI SDPKLHDNVCEKNEOFEROMSNENEMQLSKQSTFPVNDAPVCHPVD AEKINIPSDKTLVDSAGLGGGCEESAKGLNANQGRIR	341
CsueciumBBB189871-499	254	SLDQGIT EIVYKGEHNNPKPDHAKRSSGVHGOELFAQTNQDPNSD PKLHDNIC EKN E OFEROMSNENEMQLSKQSTFPVNDAPVCHPVD AEKHNPSDKTVDSAGLGGGCEESAKGLNANQGRIR	380
Cberlandier_ncafold273371-507	255	SLDQGIT EIVYKGEHNNPKPDHAKRSSGVHGOELFAQTNQDPNSD PKLHDNIC EKN E OFEROMSNENEMQLSKQSTFPVNDAPVCHPVD AEKHNPSDKTVDSAGLGGGCEESAKGLNANQGRIR	381
Quinoa_CQ001086_C68/1-507	255	SLDQGIT EIVYKGEHNNPKPDHAKRSSGVHGOELFAQTNQDPNSD PKLHDNIC EKN E OFEROMSNENEMQLSKQSTFPVNDAPVCHPVD AEKHNPSDKTVDSAGLGGGCEESAKGLNANQGRIR	381
Cberlandierinuttalliae_Cb002931_C68/1-507	255	SLDQGIT EIVYKGEHNNPKPDHAKRSSGVHGOELFAQTNQDPNSD PKLHDNIC EKN E OFEROMSNENEMQLSKQSTFPVNDAPVCHPVD AEKHNPSDKTVDSAGLGGGCEESAKGLNANQGRIR	381
Gpallidicaule_CP003660/1-465	342	NKRRKQNVQNETGKSGEGVQEP C I WVHN . EPEIMDDGFRWRKYGQKVKVGSYPRSYRCTSVKCNVRKYVERALDSDRAFITTYEGRHNHEMPKIMNLSAVLEHETAPAGKEIRTLMPIVS	465
CsueciumBBB189871-499	381	NKRRKQNVQNETGKSGEGVQEP C I WVHN I NEPEIMDDGFRWRKYGQKVKVGSYPRSYRCTSVKCNVRKYVERALDSDRAFITTYEGRHNHEMPKIMNLSAVLEHETAPAGKEIRTLMPIVS	499
Cberlandier_ncafold273371-507	382	NKRRKQNVQNETGKSGEGVQEP C I WVHN I NEPEIMDDGFRWRKYGQKVKVGSYPRSYRCTSVKCNVRKYVERALDSDRAFITTYEGRHNHEMPKIMNLSAVLEHETAPAGKEIRTLMPIVS	507
Quinoa_CQ001086_C68/1-507	382	NKRRKQNVQNETGKSGEGVQEP C I WVHN I NEPEIMDDGFRWRKYGQKVKVGSYPRSYRCTSVKCNVRKYVERALDSDRAFITTYEGRHNHEMPKIMNLSAVLEHETAPAGKEIRTLMPIVS	507
Cberlandierinuttalliae_Cb002931_C68/1-507	382	NKRRKQNVQNETGKSGEGVQEP C I WVHN I NEPEIMDDGFRWRKYGQKVKVGSYPRSYRCTSVKCNVRKYVERALDSDRAFITTYEGRHNHEMPKIMNLSAVLEHETAPAGKEIRTLMPIVS	507

I.

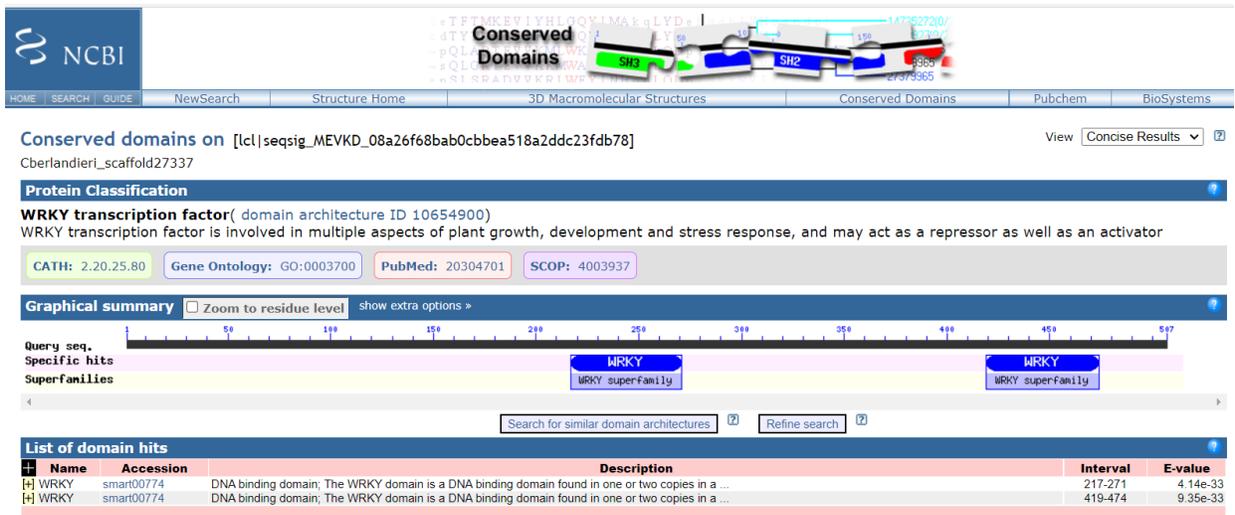


Figure S13. *TTG2* gene models. A. Potential homologs of *A. thaliana TTG2* (NP_181263) in *C. suecicum*. B. Potential homologs of *A. thaliana TTG2* (NP_181263) in *C. pallidicaule*. C. Clustal Phylogeny derived tree for potential homologs of *TTG2* in *C. pallidicaule*. D. C-terminus of *C. suecicum* homolog, showing complete open reading frame extending one amino acid further (red arrow) than the annotated protein in the database itself (n.b. reading frame goes from right to left in this window as gene is on bottom strand of the genomic scaffold as annotated). The extra C-terminal R is included in the alignment in panel E. E. Splign alignment of annotated *C. quinoa* CDS (upper aligned sequence) to genomic (lower aligned sequence) for chromosome 6A gene CQ026523, showing GC splice donor site of first intron. F. Splign alignment of *C. quinoa* chromosome 6B gene CQ001086 mRNA (upper aligned sequence) to chromosome 6A CQ026523 genomic (lower aligned sequence). Red arrows indicate positions of a 2bp insertion and a 7 bp deletion in the genomic sequence, whose net effect is to create a frameshift, severely truncating the correct open reading frame. G. Splign alignment of *C. quinoa* chromosome 6B gene CQ001086 mRNA (upper aligned sequence) to chromosome 6B sequence of *C. berlandieri nuttaliae* extended downstream of annotated 3' end of gene Cb002931; only the first exon sequence of the alignment is shown but the entire expected protein coding sequence is present across the five exons. H. Multiple sequence alignment of *TTG2* homologs, including *C. pallidicaule*, *C. suecicum*, the sub-genome B copies of *C. quinoa* and *C. berlandieri nuttaliae*, and the presumptive sub-genome B copy of the *de novo* assembly (scaffold27337). I. Predicted conserved functional domains for the *de novo* assembly scaffold27337-encoded spliced translation product.

References

1. Simon, S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. **2010**, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
2. Kokot, M.; Dlugosz, M.; Deorowicz, S. KMC 3: counting and manipulating k-mer statistics. *Bioinformatics* **2017**, *33*, 2759-2761, doi:10.1093/bioinformatics/btx304.
3. Vurture, G.W.; Sedlazeck, F.J.; Nattestad, M.; Underwood, C.J.; Fang, H.; Gurtowski, J.; Schatz, M.C. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **2017**, *33*, 2202-2204, doi:10.1093/bioinformatics/btx153.
4. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884-i890, doi:10.1093/bioinformatics/bty560.
5. Kajitani, R.; Toshimoto, K.; Noguchi, H.; Toyoda, A.; Ogura, Y.; Okuno, M.; Yabana, M.; Harada, M.; Nagayasu, E.; Maruyama, H.; et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res* **2014**, *24*, 1384-1395, doi:10.1101/gr.170720.113.
6. Jackman, S.D.; Vandervalk, B.P.; Mohamadi, H.; Chu, J.; Yeo, S.; Hammond, S.A.; Jahesh, G.; Khan, H.; Coombe, L.; Warren, R.L.; et al. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Res* **2017**, *27*, 768-777, doi:10.1101/gr.214346.116.
7. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072-1075, doi:10.1093/bioinformatics/btt086.
8. Smit, A.F.A., Hubley, R, Green, P. . RepeatMasker Open-4.0. **2015**, <http://www.repeatmasker.org>.
9. Luo, X.; Chen, S.; Zhang, Y. PlantRep: a database of plant repetitive elements. *Plant Cell Rep* **2022**, *41*, 1163-1166, doi:10.1007/s00299-021-02817-y.
10. Simao, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210-3212, doi:10.1093/bioinformatics/btv351.
11. Tarasov, A.; Vilella, A.J.; Cuppen, E.; Nijman, I.J.; Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **2015**, *31*, 2032-2034, doi:10.1093/bioinformatics/btv098.
12. Li, H. Exploring single-sample SNP and INDEL calling with whole-genome de novo assembly. *Bioinformatics* **2012**, *28*, 1838-1844, doi:10.1093/bioinformatics/bts280.
13. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve years of SAMtools and BCFtools. *Gigascience* **2021**, *10*, doi:10.1093/gigascience/giab008.
14. Lefouili, M.; Nam, K. The evaluation of Bcftools mpileup and GATK HaplotypeCaller for variant calling in non-human species. *Sci Rep* **2022**, *12*, 11331, doi:10.1038/s41598-022-15563-2.
15. Neph, S.; Kuehn, M.S.; Reynolds, A.P.; Haugen, E.; Thurman, R.E.; Johnson, A.K.; Rynes, E.; Maurano, M.T.; Vierstra, J.; Thomas, S.; et al. BEDOPS: high-performance genomic feature operations. *Bioinformatics* **2012**, *28*, 1919-1920, doi:10.1093/bioinformatics/bts277.
16. Quinlan, A.R.; Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841-842, doi:10.1093/bioinformatics/btq033.
17. Cabanettes, F.; Klopp, C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **2018**, *6*, e4958, doi:10.7717/peerj.4958.
18. Kapustin, Y.; Souvorov, A.; Tatusova, T.; Lipman, D. Splign: algorithms for computing spliced alignments with identification of paralogs. *Biol Direct* **2008**, *3*, 20, doi:10.1186/1745-6150-3-20.
19. Florea, L.; Hartzell, G.; Zhang, Z.; Rubin, G.M.; Miller, W. A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res* **1998**, *8*, 967-974, doi:10.1101/gr.8.9.967.

20. Wu, T.D.; Watanabe, C.K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **2005**, *21*, 1859-1875, doi:10.1093/bioinformatics/bti310.