

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

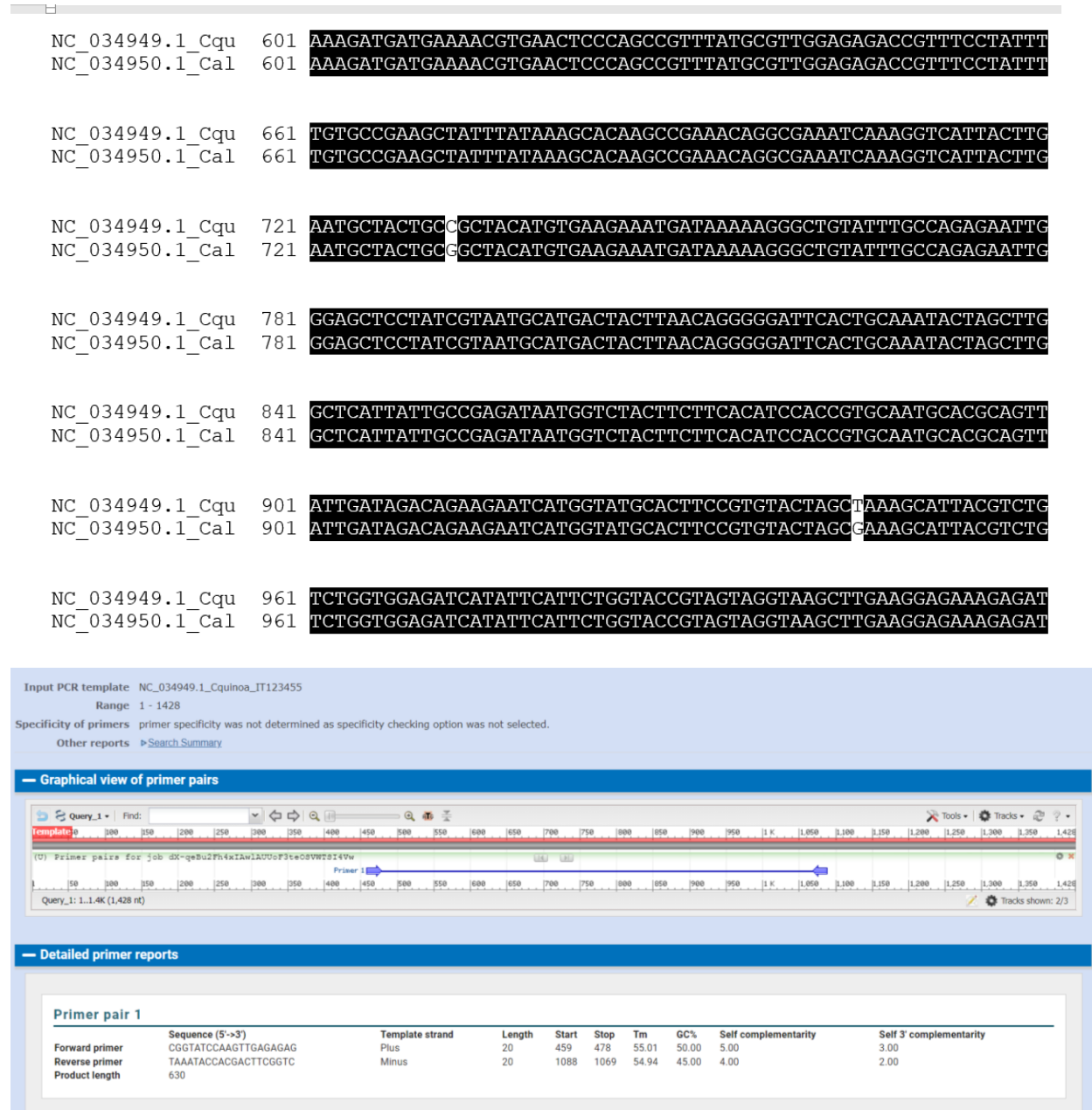


Figure S1. Primer design of *rbcl*. a, alignment of *rbcl* in *C. quinoa* and *C. album* gene as amplified by primers *rbcl*M1-F and *rbcl*M1-R, showing two sites different in the amplicon. b, PCR amplification product of *C. quinoa rbcl* using primers *rbcl*M1-F and *rbcl*M1-R, per NCBI Primer-BLAST

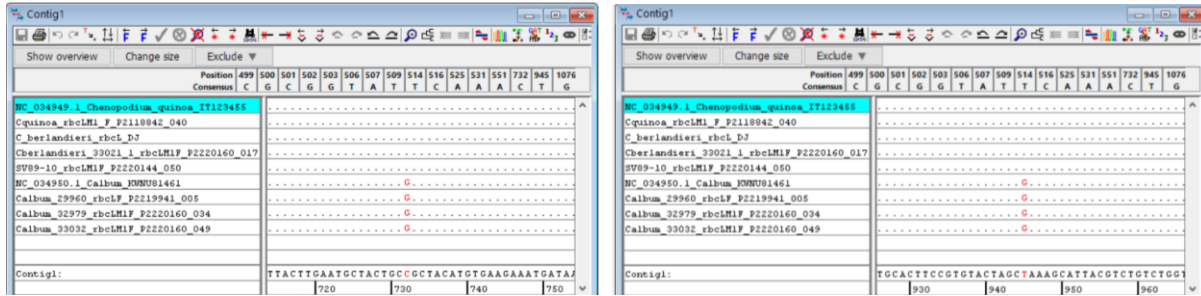


Figure S2. *rbcL* barcoding. Alignment of references and PCR-amplified DNAs, using primers *rbcLM1-F* and *rbcLM2-R*, sequenced in the forward direction using *rbcLM1-F* (identical sequences on reverse strands). Numbering is from the ATG initiation codon of the *rbcL* gene in the three references. *C. 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.

~~satK_Calypso~~ 361 AAAGAAAAAAGAAATAGTAAATCTCATAATTTGCGATCTATTCAATATTTCTT
~~satK_Calypso~~ 359 -----AAAAAGAAATAGTAAATCTCATAATTTGCGATCTATTCAATATTTCTT

~~satK_Calypso~~ 421 TTTTAGAGGACAAATCTTCCATTTCATTATGTGTAGATATATTAATAACTTACCCT
~~satK_Calypso~~ 412 TTTTAGAGGACAAATCTTCCATTTCATTATGTGTAGATATATTAATAACTTACCCT

~~satK_Calypso~~ 481 GCCCATCTAGAAATCTTGGTTCAAACTCTTCGCTACTGGTTGAAAGATGCCTCTTCTTTG
~~satK_Calypso~~ 472 GCCCATCTAGAAATCTTGGTTCAAACTCTTCGCTACTGGTTGAAAGATGCCTCTTCTTTG

~~satK_Calypso~~ 541 CATTTAGTAAGATTTTTCTTTATGAGTCTCGTAATTTAAATAGTCTTATTATCCAAAG
~~satK_Calypso~~ 532 CATTTAGTAAGATTTTTCTTTATGAGTCTCGTAATTTAAATAGTCTTATTATCCAAAG

~~satK_Calypso~~ 601 GAATTCATTTCTTTTGAAGAAAGGAATCAAGATTATTCTTGTTCCTATATAATTTG
~~satK_Calypso~~ 592 GAATTCATTTCTTTTGAAGAAAGGAATCAAGATTATTCTTGTTCCTATATAATTTG

~~satK_Calypso~~ 661 CATGTATGCGAATACGAATCTTTTGTGTTTTCTCCGCAACCAATCCTCTTATTACGA
~~satK_Calypso~~ 652 CATGTATGCGAATACGAATCTTTTGTGTTTTCTCCGCAACCAATCCTCTTATTACGA

~~satK_Calypso~~ 721 TCACCTCTTTTGGAAACCTTATTGAACGAATTTTTCTACGGAAGAACTAAAAATATCTA
~~satK_Calypso~~ 712 TCACCTCTTTTGGAAACCTTATTGAACGAATTTTTCTACGGAAGAACTAAAAATATCTA

~~satK_Calypso~~ 781 GTAAAGTTTTTACTAAGGATTTTGGGGTTATCCTATGGCTTTTCAGAGAACCTTTCCCG
~~satK_Calypso~~ 772 GTAAAGTTTTTACTAAGGATTTTGGGGTTATCCTATGGCTTTTCAGAGAACCTTTCCCG

~~satK_Calypso~~ 841 CATTATGTTAGGTATCAAGGAAATCAATTCTAGCCTCAAGAGGGACATCTCTTTTGATG
~~satK_Calypso~~ 832 CATTATGTTAGGTATCAAGGAAATCAATTCTAGCCTCAAGAGGGACATCTCTTTTGATG

~~satK_Calypso~~ 901 CATAAATGGAAAAATTACCTTATCCATTCTGGCAATGTAATTTTCTGTGTGGTCTCAA
~~satK_Calypso~~ 892 CATAAATGGAAAAATTACCTTATCCATTCTGGCAATGTAATTTTCTGTGTGGTCTCAA

~~satK_Calypso~~ 961 CCAGAAAGCTCTATATCAATCGATTATCAAGCAATTCCTTAGACTTTATGGGTTTTCTT
~~satK_Calypso~~ 952 CCAGAAAGCTCTATATCAATCGATTATCAAGCAATTCCTTAGACTTTATGGGTTTTCTT

~~satK_Calypso~~ 1021 TCAGATGTGCGACTCAATTCTTCAGTAATACGGAGTCAATGTTGAGAAATGCATTCTTA
~~satK_Calypso~~ 1012 TCAGATGTGCGACTCAATTCTTCAGTAATACGGAGTCAATGTTGAGAAATGCATTCTTA

~~satK_Calypso~~ 1081 ATAGAGAAATATTATTAAGAAATTCGATGCCACAGTCCCAATTATTCCTCTGGTTGGATCG
~~satK_Calypso~~ 1072 ATAGAGAAATATTATTAAGAAATTCGATGCCACAGTCCCAATTATTCCTCTGGTTGGATCG

~~satK_Calypso~~ 1141 TTGGCTAAAGCGAAATTTTGTAAATGATTAGGGCATCCCGTTAGTAAGTCGGTCTGGACC
~~satK_Calypso~~ 1132 TTGGCTAAAGCGAAATTTTGTAAATGATTAGGGCATCCCGTTAGTAAGTCGGTCTGGACC

~~satK_Calypso~~ 1201 GATTATCCGATTCTGATATTATTGATCGATTGGACGTATCTGTAGAACTATTTTTCAT
~~satK_Calypso~~ 1192 GATTATCCGATTCTGATATTATTGATCGATTGGACGTATCTGTAGAACTATTTTTCAT

~~satK_Calypso~~ 1261 TATTATAGTGGCTCTTCAGAAAAAAGTTTGTATCGAATAAAGTATATCTTCGATT
~~satK_Calypso~~ 1252 TATTATAGTGGCTCTTCAGAAAAAAGTTTGTATCGAATAAAGTATATCTTCGATT

~~satK_Calypso~~ 1321 TCTGTGCTAGAACTTTGTCTCGTAAACACAAAGTACTGTACGTGCTTTTTGAAAAGA
~~satK_Calypso~~ 1312 TCTGTGCTAGAACTTTGTCTCGTAAACACAAAGTACTGTACGTGCTTTTTGAAAAGA



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.

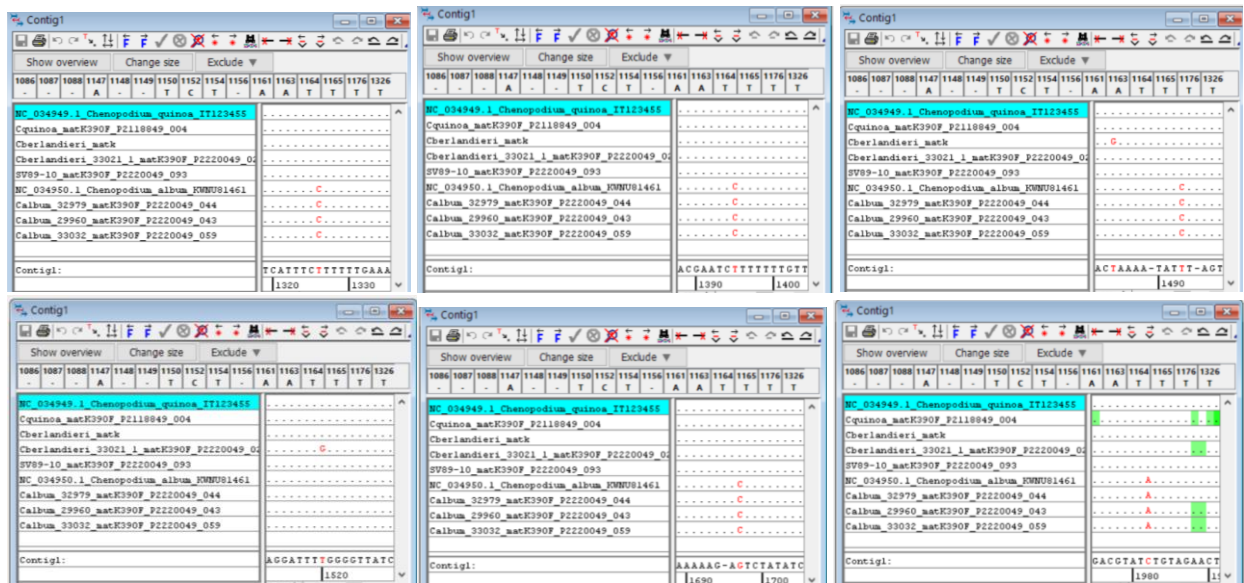
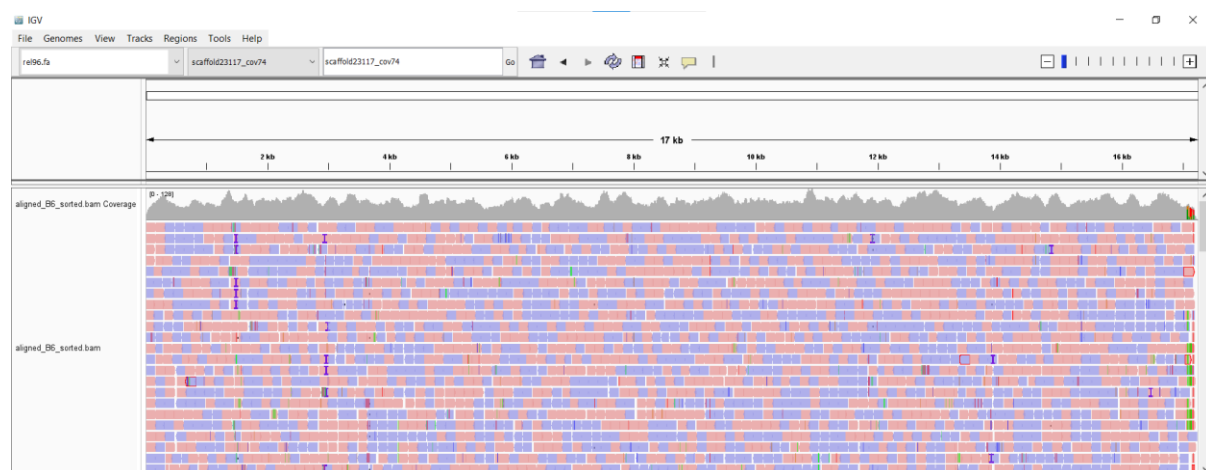


Figure S4. *matK* barcoding. Alignment of references and PCR-amplified DNAs, using primers matK390-F and matK3FKIM-R, sequenced in the forward direction using matK390-F (identical sequences on reverse strands). Numbering is from a position in the *C. berlandieri* chloroplast genome approximately 710 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.

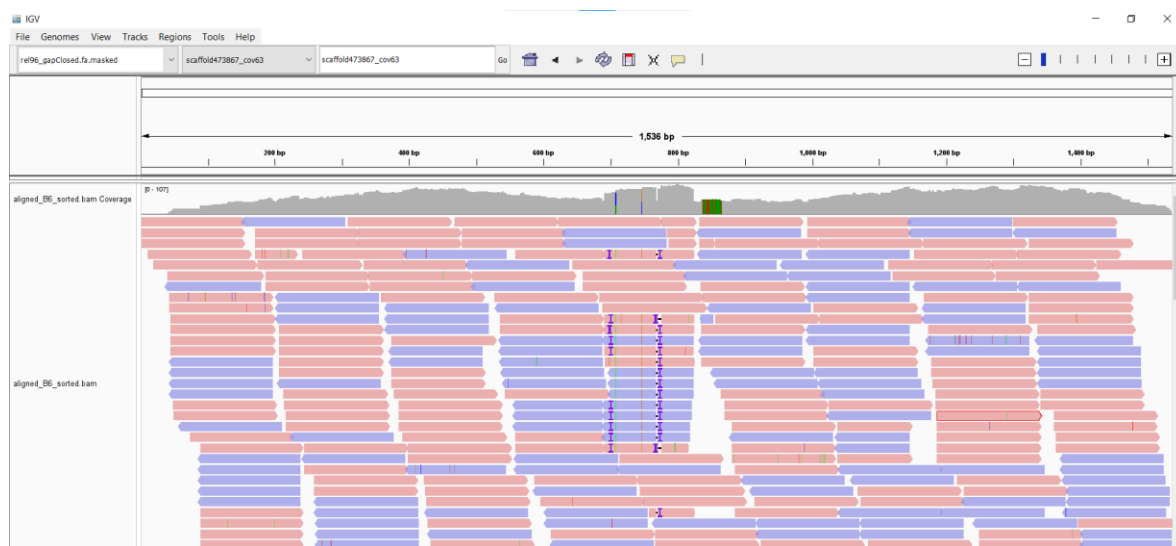
A.



B.



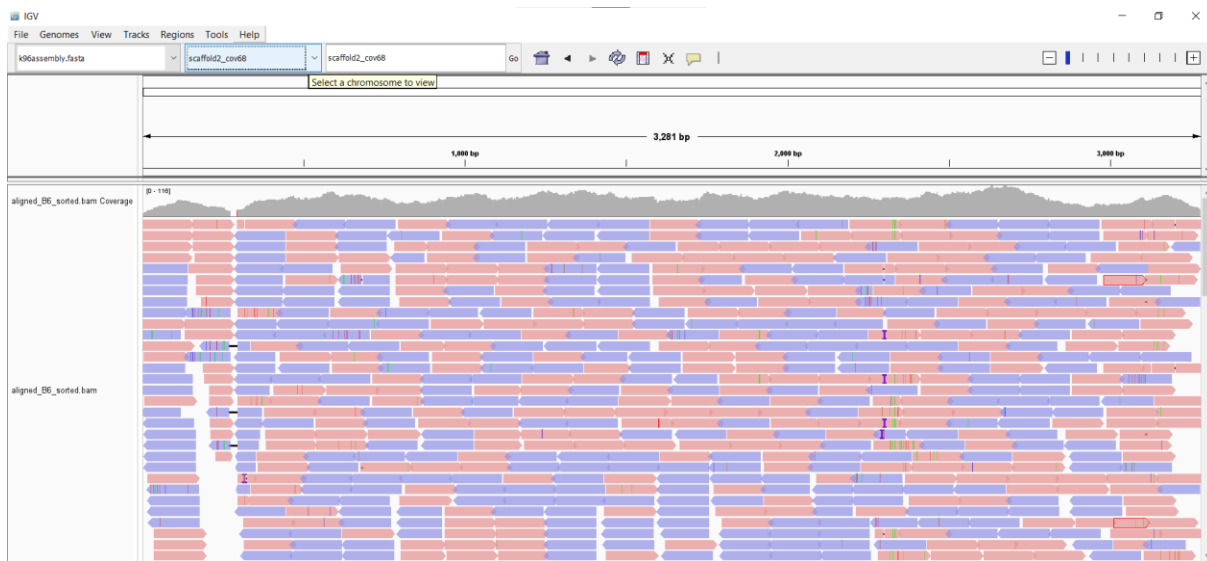
C.



D.



E.



F.

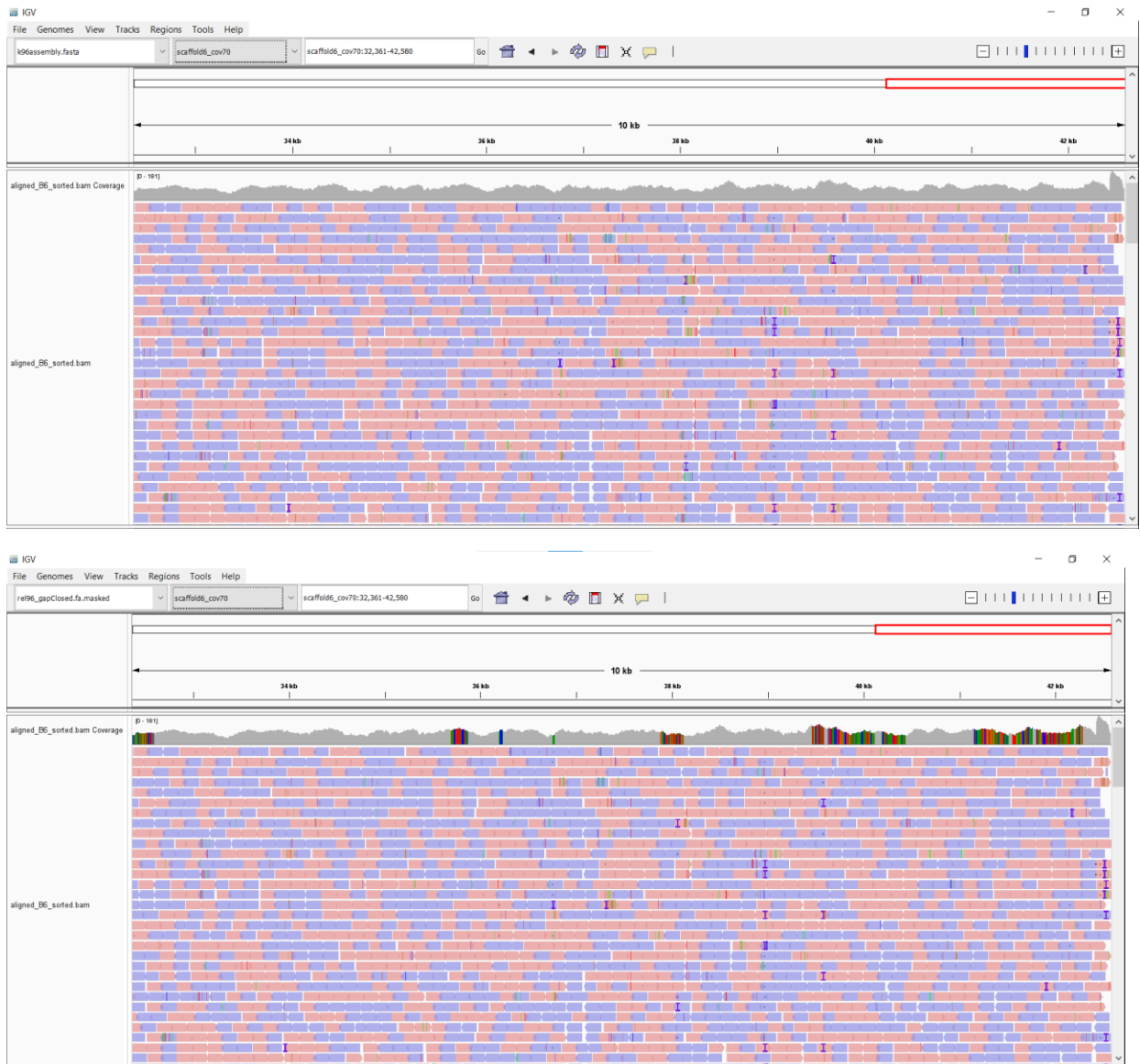
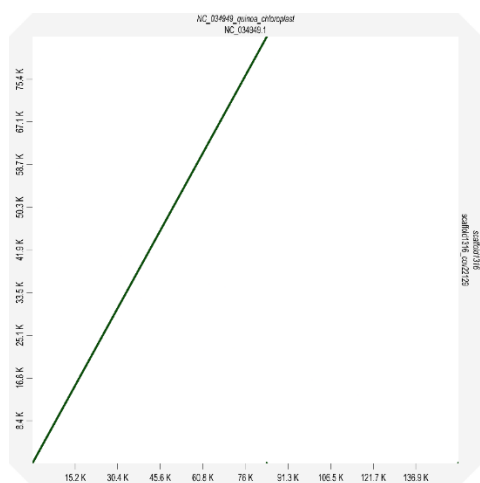
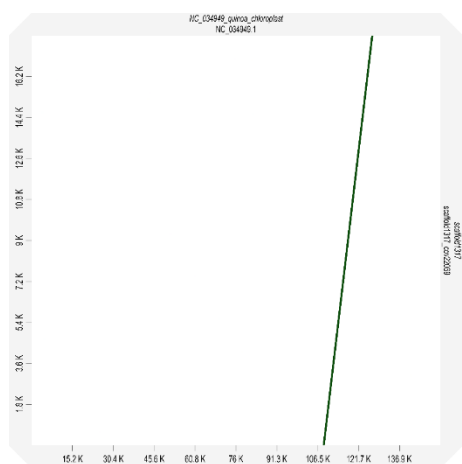


Figure S5. Selected scaffolds showing examples of heterozygous SNPs. Each scaffold is shown twice, first aligned to the whole genome assembly, then below aligned to the genome assembly with repeats masked. Split vertical colored lines present in both alignments of the scaffold represent heterozygous 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.

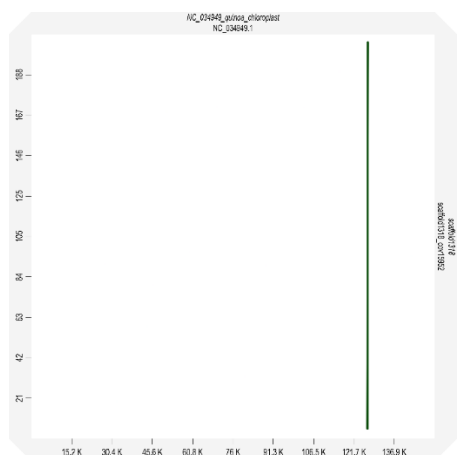
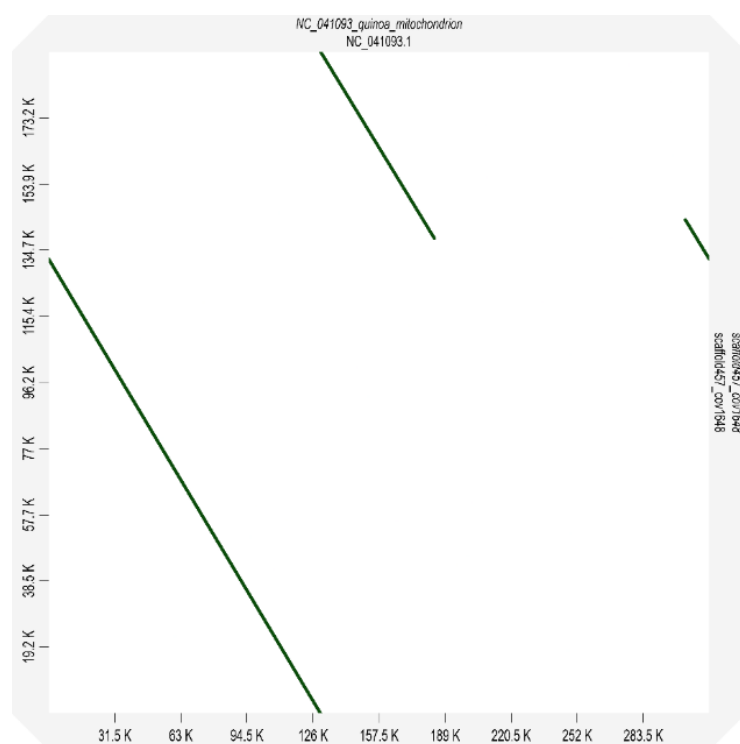


Figure S6. Chloroplast genome assembly. A, B, C, alignments to *C. quinoa* unique chloroplast genome regions.

A.



B.

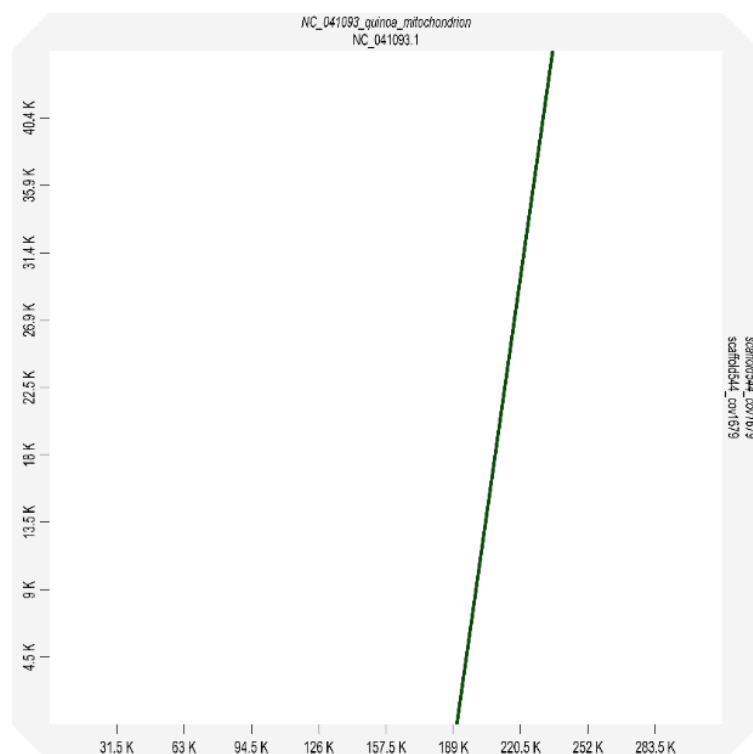


Figure S7. Mitochondrial genome assembly. A, B, alignments to *C. quinoa* mitochondrial genome.

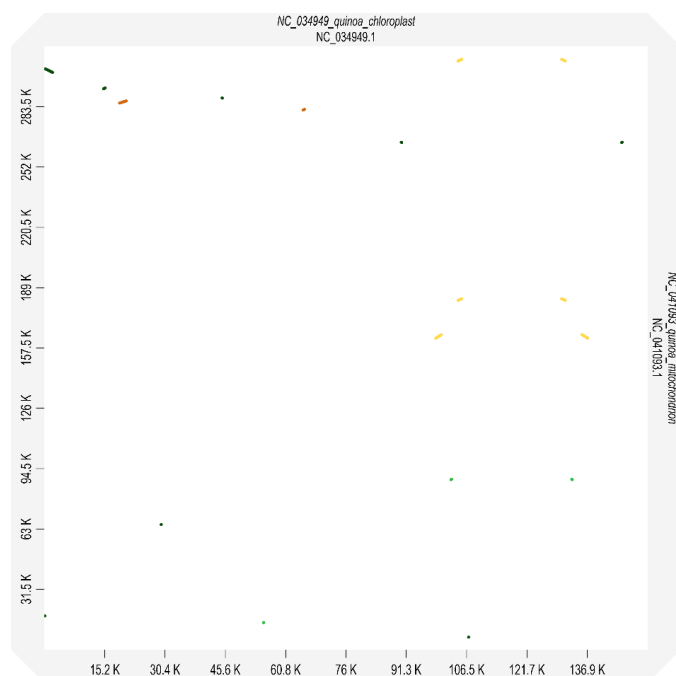


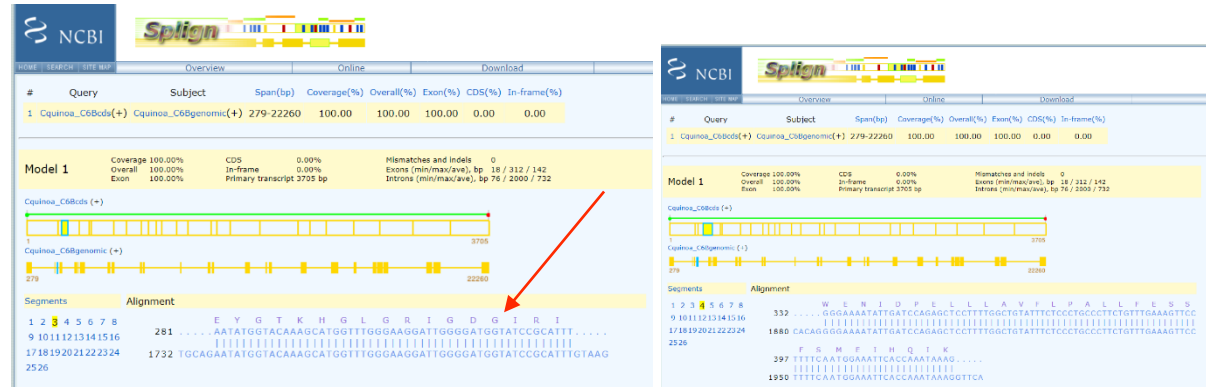
Figure S8. Alignment of *C. quinoa* chloroplast and mitochondrial genomes to each other.

Samuels Genomic sequence of Canadian *Chenopodium berlandieri*

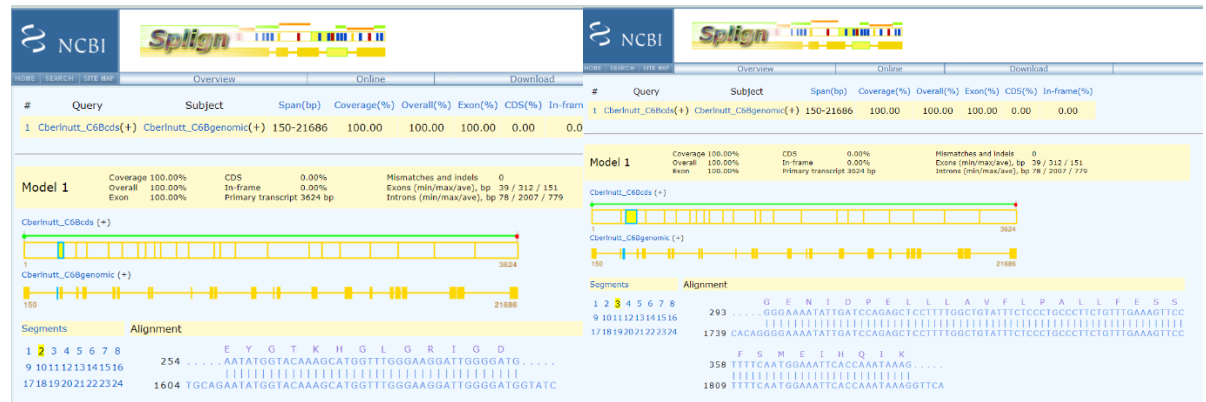
A.

Quinoa_S051_C6B1-1234	1	MAAYRIAVPFLMLAEVGAETTTPTNSTISGSAMAESESTPTDAVIFGVSLILGIACRHLRGTTRVPTVALLIGIGLGLSDSEVITSVEYQTKHGLRIGDGI	150
Coelebre_S051-1-1211	1	MAAYRIAVPFLMLAEVGAETTTPTNSTISGSAMAESESTPTDAVIFGVSLILGIACRHLRGTTRVPTVALLIGIGLGLSDSEVITSVEYQTKHGLRIGDGI	141
Cberlandierutallae_S051_C6B1-1207	1	MAAYRIAVPFLMLAEVGAETTTPTNSTISGSAMAESESTPTDAVIFGVSLILGIACRHLRGTTRVPTVALLIGIGLGLSDSEVITSVEYQTKHGLRIGDGI	137
SVB9-10_S051_C6B1-1207	1	MAAYRIAVPFLMLAEVGAETTTPTNSTISGSAMAESESTPTDAVIFGVSLILGIACRHLRGTTRVPTVALLIGIGLGLSDSEVITSVEYQTKHGLRIGDGI	137
Quinoa_S051_C6B1-1234	161	AGPVLVISTFCLGAALKWIDKALSFYDWSWKTSLLLGGLLSATDPVAVALLKELGASKSLSTIEGESLMDGDTLSLNTAIVVYQLFKMLLORTFNWASILKYLVGVTFGAVGFIAGFASVWLWGFIFNDTVIEITLTLAV	300
Coelebre_S051-1-1211	142	AGPVLVISTFCLGAALKWIDKALSFYDWSWKTSLLLGGLLSATDPVAVALLKELGASKSLSTIEGESLMDGDTLSLNTAIVVYQLFKMLLORTFNWASILKYLVGVTFGAVGFIAGFASVWLWGFIFNDTVIEITLTLAV	277
Cberlandierutallae_S051_C6B1-1207	138	AGPVLVISTFCLGAALKWIDKALSFYDWSWKTSLLLGGLLSATDPVAVALLKELGASKSLSTIEGESLMDGDTLSLNTAIVVYQLFKMLLORTFNWASILKYLVGVTFGAVGFIAGFASVWLWGFIFNDTVIEITLTLAV	273
SVB9-10_S051_C6B1-1207	138	AGPVLVISTFCLGAALKWIDKALSFYDWSWKTSLLLGGLLSATDPVAVALLKELGASKSLSTIEGESLMDGDTLSLNTAIVVYQLFKMLLORTFNWASILKYLVGVTFGAVGFIAGFASVWLWGFIFNDTVIEITLTLAV	273
Quinoa_S051_C6B1-1234	301	SVYAVFTAGEGADVSGVLVTMLGMFYAAARATFQESQSLSHHFHEMVAIVANTLIFILSGAVIAGVLSNDIFENHONASVSKLSVSNYSMEDVNLGIEKWIQHTLGRLSLSDVGVVLFELDFDGTAWGVLLLVVYVLVARGV	400
Coelebre_S051-1-1211	278	SVYAVFTAGEGADVSGVLVTMLGMFYAAARATFQESQSLSHHFHEMVAIVANTLIFILSGAVIAGVLSNDIFENHONASVSKLSVSNYSMEDVNLGIEKWIQHTLGRLSLSDVGVVLFELDFDGTAWGVLLLVVYVLVARGV	427
Cberlandierutallae_S051_C6B1-1207	274	SVYAVFTAGEGADVSGVLVTMLGMFYAAARATFQESQSLSHHFHEMVAIVANTLIFILSGAVIAGVLSNDIFENHONASVSKLSVSNYSMEDVNLGIEKWIQHTLGRLSLSDVGVVLFELDFDGTAWGVLLLVVYVLVARGV	423
SVB9-10_S051_C6B1-1207	274	SVYAVFTAGEGADVSGVLVTMLGMFYAAARATFQESQSLSHHFHEMVAIVANTLIFILSGAVIAGVLSNDIFENHONASVSKLSVSNYSMEDVNLGIEKWIQHTLGRLSLSDVGVVLFELDFDGTAWGVLLLVVYVLVARGV	423
Quinoa_S051_C6B1-1234	451	VGVVLYPFLCYFQYQMEWKEAMILWAGLRGAVALSLSLSVKRSQDPAYLSTGTGTFLVFFTTGIVFLTLINOSTTQVFLRLGMDLKSARKRILEFTKYEMKKALAEAFQDLGDELOPADMTVKRYIKSLNTISDRIHPHDA	600
Coelebre_S051-1-1211	428	VGVVLYPFLCYFQYQMEWKEAMILWAGLRGAVALSLSLSVKRSQDPAYLSTGTGTFLVFFTTGIVFLTLINOSTTQVFLRLGMDLKSARKRILEFTKYEMKKALAEAFQDLGDELOPADMTVKRYIKSLNTISDRIHPHDA	577
Cberlandierutallae_S051_C6B1-1207	424	VGVVLYPFLCYFQYQMEWKEAMILWAGLRGAVALSLSLSVKRSQDPAYLSTGTGTFLVFFTTGIVFLTLINOSTTQVFLRLGMDLKSARKRILEFTKYEMKKALAEAFQDLGDELOPADMTVKRYIKSLNTISDRIHPHDA	573
SVB9-10_S051_C6B1-1207	424	VGVVLYPFLCYFQYQMEWKEAMILWAGLRGAVALSLSLSVKRSQDPAYLSTGTGTFLVFFTTGIVFLTLINOSTTQVFLRLGMDLKSARKRILEFTKYEMKKALAEAFQDLGDELOPADMTVKRYIKSLNTISDRIHPHDA	573
Quinoa_S051_C6B1-1234	601	SDTSDNGLDPMNKLDMRVLNLGVQSAAYWMLDEGRITOSTANVLMSVDEALDAVDHPELGDWKLKNSVHPFKYRLLGGLVYPKKLVTFFTVERLESACYICAAFLRAHRTARGOLHDFIGDSEISAVITESETEGEARKFL	720
Coelebre_S051-1-1211	575	SDTSDNGLDPMNKLDMRVLNLGVQSAAYWMLDEGRITOSTANVLMSVDEALDAVDHPELGDWKLKNSVHPFKYRLLGGLVYPKKLVTFFTVERLESACYICAAFLRAHRTARGOLHDFIGDSEISAVITESETEGEARKFL	727
Cberlandierutallae_S051_C6B1-1207	574	SDTSDNGLDPMNKLDMRVLNLGVQSAAYWMLDEGRITOSTANVLMSVDEALDAVDHPELGDWKLKNSVHPFKYRLLGGLVYPKKLVTFFTVERLESACYICAAFLRAHRTARGOLHDFIGDSEISAVITESETEGEARKFL	723
SVB9-10_S051_C6B1-1207	574	SDTSDNGLDPMNKLDMRVLNLGVQSAAYWMLDEGRITOSTANVLMSVDEALDAVDHPELGDWKLKNSVHPFKYRLLGGLVYPKKLVTFFTVERLESACYICAAFLRAHRTARGOLHDFIGDSEISAVITESETEGEARKFL	723
Quinoa_S051_C6B1-1234	751	VRTTFPEVLRRVKTQVTVAYVLQHLIEYIESLEKAGLLEKEMLHLHDAVDTDLKRLVRNPPTVKIKIGELISMHPGLALPSGVRDLVSTKEEVKVRMTLYKEGQKPNGLISNOGVKWSKVRKKNHALHOTFTHSTLGLYE	900
Coelebre_S051-1-1211	728	VRTTFPEVLRRVKTQVTVAYVLQHLIEYIESLEKAGLLEKEMLHLHDAVDTDLKRLVRNPPTVKIKIGELISMHPGLALPSGVRDLVSTKEEVKVRMTLYKEGQKPNGLISNOGVKWSKVRKKNHALHOTFTHSTLGLYE	877
Cberlandierutallae_S051_C6B1-1207	724	VRTTFPEVLRRVKTQVTVAYVLQHLIEYIESLEKAGLLEKEMLHLHDAVDTDLKRLVRNPPTVKIKIGELISMHPGLALPSGVRDLVSTKEEVKVRMTLYKEGQKPNGLISNOGVKWSKVRKKNHALHOTFTHSTLGLYE	873
SVB9-10_S051_C6B1-1207	724	VRTTFPEVLRRVKTQVTVAYVLQHLIEYIESLEKAGLLEKEMLHLHDAVDTDLKRLVRNPPTVKIKIGELISMHPGLALPSGVRDLVSTKEEVKVRMTLYKEGQKPNGLISNOGVKWSKVRKKNHALHOTFTHSTLGLYE	873
Quinoa_S051_C6B1-1234	1001	VLIOKPYLMDITDSVAVCFYIETEKILAAAGSDPAVEHFWKESIVLAKVLLPVRFEMSDMDMLKTAERSTLNTYLRGTEIEVPSHSIGFLLEQIKSHSLVEELITSPALWPAQNSLFLSDEGQYKSTFSLHOGASYVYETR	1050
Coelebre_S051-1-1211	978	VLIOKPYLMDITDSVAVCFYIETEKILAAAGSDPAVEHFWKESIVLAKVLLPVRFEMSDMDMLKTAERSTLNTYLRGTEIEVPSHSIGFLLEQIKSHSLVEELITSPALWPAQNSLFLSDEGQYKSTFSLHOGASYVYETR	1027
Cberlandierutallae_S051_C6B1-1207	974	VLIOKPYLMDITDSVAVCFYIETEKILAAAGSDPAVEHFWKESIVLAKVLLPVRFEMSDMDMLKTAERSTLNTYLRGTEIEVPSHSIGFLLEQIKSHSLVEELITSPALWPAQNSLFLSDEGQYKSTFSLHOGASYVYETR	1023
SVB9-10_S051_C6B1-1207	974	VLIOKPYLMDITDSVAVCFYIETEKILAAAGSDPAVEHFWKESIVLAKVLLPVRFEMSDMDMLKTAERSTLNTYLRGTEIEVPSHSIGFLLEQIKSHSLVEELITSPALWPAQNSLFLSDEGQYKSTFSLHOGASYVYETR	1023
Quinoa_S051_C6B1-1234	1051	ARVLIDMVPIDAGNTLLRRKSLLHDSRSRLNSRDHAGLLSWPENQYKSHRPLDGOEIGDSOKLSAKAMLEIYOSTARDVPLRLGSLFOGYSLONPSHVRSPQVPIGOKRPLTSVKSEBSNTVRKLGEQVMREELLPTQTHSRH	1200
Coelebre_S051-1-1211	1028	ARVLIDMVPIDAGNTLLRRKSLLHDSRSRLNSRDHAGLLSWPENQYKSHRPLDGOEIGDSOKLSAKAMLEIYOSTARDVPLRLGSLFOGYSLONPSHVRSPQVPIGOKRPLTSVKSEBSNTVRKLGEQVMREELLPTQTHSRH	1177
Cberlandierutallae_S051_C6B1-1207	1024	ARVLIDMVPIDAGNTLLRRKSLLHDSRSRLNSRDHAGLLSWPENQYKSHRPLDGOEIGDSOKLSAKAMLEIYOSTARDVPLRLGSLFOGYSLONPSHVRSPQVPIGOKRPLTSVKSEBSNTVRKLGEQVMREELLPTQTHSRH	1173
SVB9-10_S051_C6B1-1207	1024	ARVLIDMVPIDAGNTLLRRKSLLHDSRSRLNSRDHAGLLSWPENQYKSHRPLDGOEIGDSOKLSAKAMLEIYOSTARDVPLRLGSLFOGYSLONPSHVRSPQVPIGOKRPLTSVKSEBSNTVRKLGEQVMREELLPTQTHSRH	1173
Quinoa_S051_C6B1-1234	1201	PSRVVDDSSSSSGGDEIVIRIDPSKLSFROAP	1234
Coelebre_S051-1-1211	1178	PSRVVDDSSSSSGGDEIVIRIDPSKLSFROAP	1211
Cberlandierutallae_S051_C6B1-1207	1174	PSRVVDDSSSSSGGDEIVIRIDPSKLSFROAP	1207
SVB9-10_S051_C6B1-1207	1174	PSRVVDDSSSSSGGDEIVIRIDPSKLSFROAP	1207

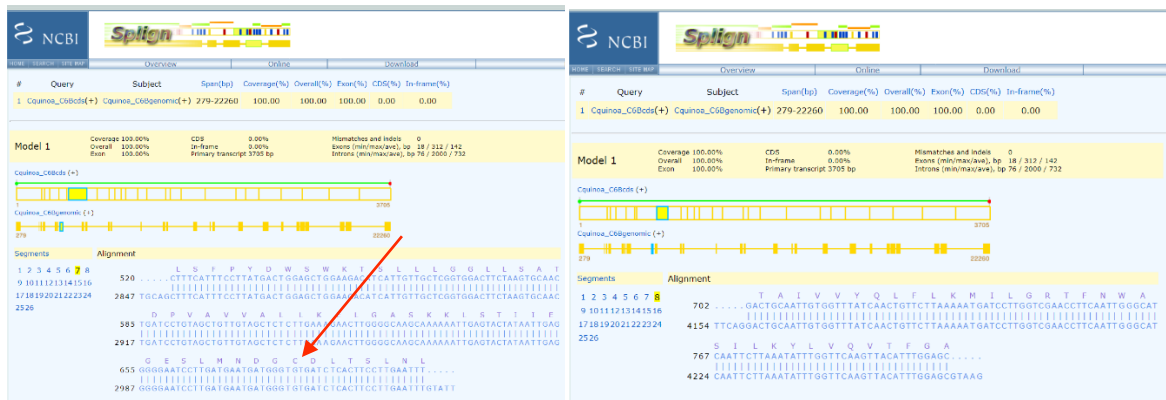
B.



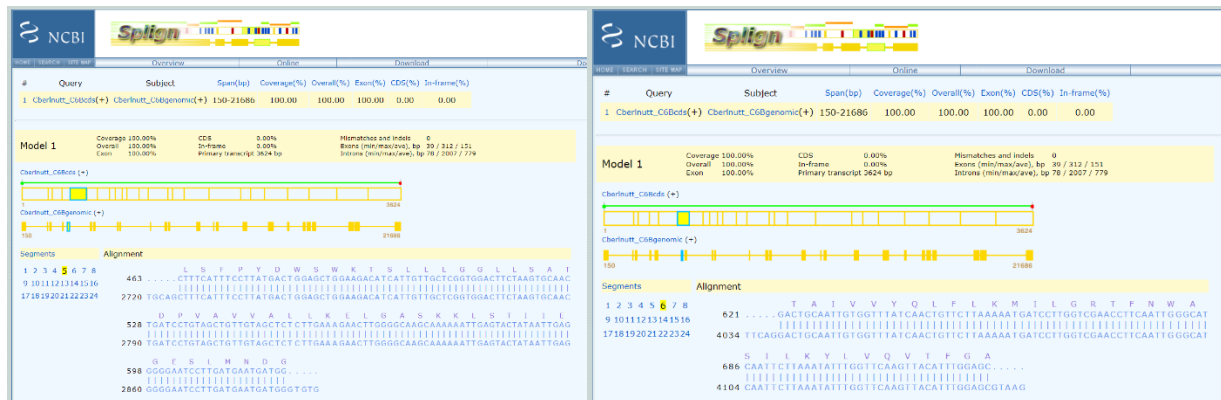
C.



D.



E.



F.

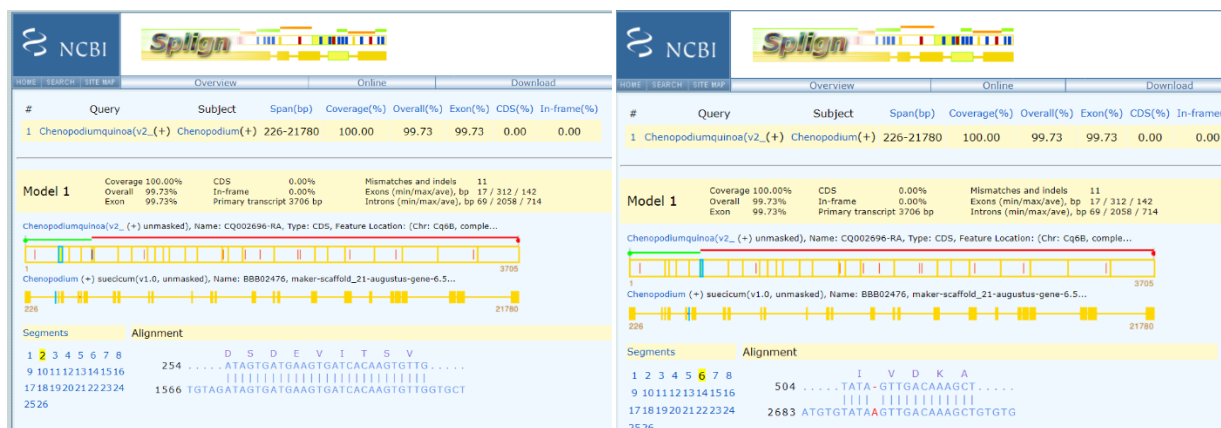


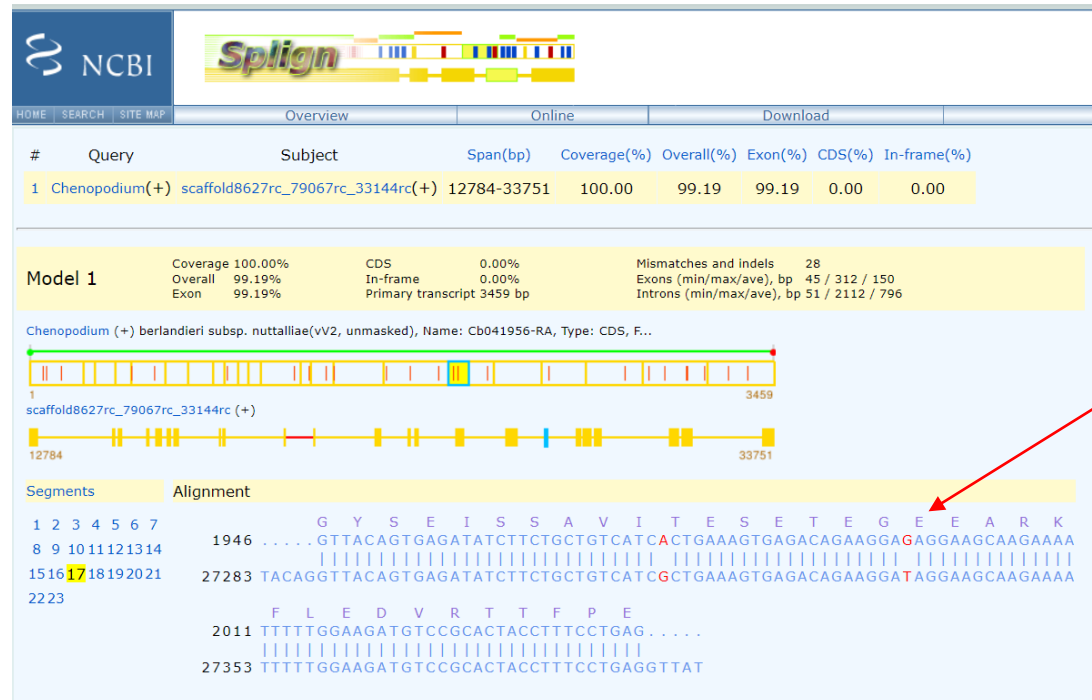
Figure S9. Gene models of *SOS1* sub-genome B. In all panels, the sequence whose name begins SV89-10 is from the *de novo* assembly reported here. A. Multiple sequence alignment of gene models from CoGe database for *C. quinoa* sub-genome B, *C. berlandieri nuttaliae*, and *C. suecicum* (the sub-genome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

A.

Quinoa_S051_C6A1-1281	1	MAAYRIAVPFLMLAEVGAETTPMNSTISASAMAEESNPTDAVIFFGVSLIGIACRHFLOTRVPYTVALLIIGIOLGSLSDSEVITSVEYOTKHOLRFDDGIRIWIENIDPELLAVLPALLFESSFSMEIHOIKRCAAGMILL	150
Quinoa_S051_C6A1-1211	1	MAAYRIAVPFLMLAEVGAETTPMNSTISASAMAEESNPTDAVIFFGVSLIGIACRHFLOTRVPYTVALLIIGIOLGSLSDSEVITSVEYOTKHOLRFDDGIRIWIENIDPELLAVLPALLFESSFSMEIHOIKRCAAGMILL	141
Chenut_S051_C6A1-1152	1	MAAYRIAVPFLMLAEVGAETTPMNSTISASAMAEESNPTDAVIFFGVSLIGIACRHFLOTRVPYTVALLIIGIOLGSLSDSEVITSVEYOTKHOLRFDDGIRIWIENIDPELLAVLPALLFESSFSMEIHOIKRCAAGMILL	141
SVB9-10_S051_C6A0000001-665	1	MAAYRIAVPFLMLAEVGAETTPMNSTISASAMAEESNPTDAVIFFGVSLIGIACRHFLOTRVPYTVALLIIGIOLGSLSDSEVITSVEYOTKHOLRFDDGIRIWIENIDPELLAVLPALLFESSFSMEIHOIKRCAAGMILL	141
Quinoa_S051_C6A1-1281	151	AGPQVLSTFLGAAALVIVVQVLSFFYDWSWKTSLLOGLLSATDPVAVALLKELGASKKLSIIIEGESLMNDGDLSTLNLVSMFLAEYICFVRTIVTAIVVYQLFKMLLORTFNWASILKYLQVQSFQAVGFIATGIASVWLW	300
Quinoa_S051_C6A1-1211	142	AGPQVLSTFLGAAALVIVVQVLSFFYDWSWKTSLLOGLLSATDPVAVALLKELGASKKLSIIIEGESLMNDGDLSTLNLVSMFLAEYICFVRTIVTAIVVYQLFKMLLORTFNWASILKYLQVQSFQAVGFIATGIASVWLW	295
Chenut_S051_C6A1-1152	142	AGPQVLSTFLGAAALVIVVQVLSFFYDWSWKTSLLOGLLSATDPVAVALLKELGASKKLSIIIEGESLMNDGDLSTLNLVSMFLAEYICFVRTIVTAIVVYQLFKMLLORTFNWASILKYLQVQSFQAVGFIATGIASVWLW	295
SVB9-10_S051_C6A0000001-665	142	AGPQVLSTFLGAAALVIVVQVLSFFYDWSWKTSLLOGLLSATDPVAVALLKELGASKKLSIIIEGESLMNDGDLSTLNLVSMFLAEYICFVRTIVTAIVVYQLFKMLLORTFNWASILKYLQVQSFQAVGFIATGIASVWLW	295
Quinoa_S051_C6A1-1281	301	GFIFNDTIIETLTLAVSYAAYFTVIWKRDRAGSAGSEKQIRTKRRGIMRENEEDRKISSVMAQEGADVSSVLTMTLMGFYAAARTAFKQESQDLSLHFWEMVAYIANTLIFILSGAVIAGVLSNDNIFENHGNASVSKLVNSV	450
Quinoa_S051_C6A1-1211	280	GFIFNDTIIETLTLAVSYAAYFTVIWKRDRAGSAGSEKQIRTKRRGIMRENEEDRKISSVMAQEGADVSSVLTMTLMGFYAAARTAFKQESQDLSLHFWEMVAYIANTLIFILSGAVIAGVLSNDNIFENHGNASVSKLVNSV	380
Chenut_S051_C6A1-1152	281	GFIFNDTIIETLTLAVSYAAYFTVIWKRDRAGSAGSEKQIRTKRRGIMRENEEDRKISSVMAQEGADVSSVLTMTLMGFYAAARTAFKQESQDLSLHFWEMVAYIANTLIFILSGAVIAGVLSNDNIFENHGNASVSKLVNSV	387
SVB9-10_S051_C6A0000001-665	281	GFIFNDTIIETLTLAVSYAAYFTVIWKRDRAGSAGSEKQIRTKRRGIMRENEEDRKISSVMAQEGADVSSVLTMTLMGFYAAARTAFKQESQDLSLHFWEMVAYIANTLIFILSGAVIAGVLSNDNIFENHGNASVSKLVNSV	387
Quinoa_S051_C6A1-1281	451	SMYKLNLSYLMIAAGHNR...EDHNESEEMHGTAWGYLLVYVVLVARGVGVVLYPFLCYFGYQMEWKEAMILVWAGLRQAVALSLSLKVRSDDPAYLSTGTGTLVFFVTGIVFLTLINOSTTQVQLFLQMDKLSKAKRR	597
Quinoa_S051_C6A1-1211	381	QDQLEKRIQ...LHTLQHLSDLSGVVFFELDSGTAWGYLLVYVVLVARGVGVVLYPFLCYFGYQMEWKEAMILVWAGLRQAVALSLSLKVRSDDPAYLSTGTGTLVFFVTGIVFLTLINOSTTQVQLFLQMDKLSKAKRR	528
Chenut_S051_C6A1-1152	358	...GTAWGYLLVYVVLVARGVGVVLYPFLCYFGYQMEWKEAMILVWAGLRQAVALSLSLKVRSDDPAYLSTGTGTLVFFVTGIVFLTLINOSTTQVQLFLQMDKLSKAKRR	474
SVB9-10_S051_C6A0000001-665	358	...GTAWGYLLVYVVLVARGVGVVLYPFLCYFGYQMEWKEAMILVWAGLRQAVALSLSLKVRSDDPAYLSTGTGTLVFFVTGIVFLTLINOSTTQVQLFLQMDKLSKAKRR	474
Quinoa_S051_C6A1-1281	598	LEFTKYEMKKALEAFODLGEDELQADWPVTVKRYIKSLNSIDGDRIPHDPASDNGFLDPMNLKDMRVLLNGVQAAWMLDEGRITOSTANVLMQSDVDEALDSVDHEFLCDWGLKNSVHFPHYRLDQGIYPKKLVTFITVERLE	747
Quinoa_S051_C6A1-1211	529	LEFTKYEMKKALEAFODLGEDELQADWPVTVKRYIKSLNSIDGDRIPHDPASDNGFLDPMNLKDMRVLLNGVQAAWMLDEGRITOSTANVLMQSDVDEALDSVDHEFLCDWGLKNSVHFPHYRLDQGIYPKKLVTFITVERLE	678
Chenut_S051_C6A1-1152	475	LEFTKYEMKKALEAFODLGEDELQADWPVTVKRYIKSLNSIDGDRIPHDPASDNGFLDPMNLKDMRVLLNGVQAAWMLDEGRITOSTANVLMQSDVDEALDSVDHEFLCDWGLKNSVHFPHYRLDQGIYPKKLVTFITVERLE	624
SVB9-10_S051_C6A0000001-665	475	LEFTKYEMKKALEAFODLGEDELQADWPVTVKRYIKSLNSIDGDRIPHDPASDNGFLDPMNLKDMRVLLNGVQAAWMLDEGRITOSTANVLMQSDVDEALDSVDHEFLCDWGLKNSVHFPHYRLDQGIYPKKLVTFITVERLE	624
Quinoa_S051_C6A1-1281	748	SACYICAALRAHRTARGLHDFIOYSEISSAVITESTETEGEARKFLEDRVTITPEVLRVKTRQVTVAYVLOHIEYIESLEKAGILEEKMHLHDAVQDGLKRLVRNPPTVKIKIGELISMHPFLGALPQSVGRDLVQSTKEEVK	897
Quinoa_S051_C6A1-1211	679	SACYICAALRAHRTARGLHDFIOYSEISSAVITESTETEGEARKFLEDRVTITPEVLRVKTRQVTVAYVLOHIEYIESLEKAGILEEKMHLHDAVQDGLKRLVRNPPTVKIKIGELISMHPFLGALPQSVGRDLVQSTKEEVK	628
Chenut_S051_C6A1-1152	625	SACYICAALRAHRTARGLHDFIOYSEISSAVITESTETEGEARKFLEDRVTITPEVLRVKTRQVTVAYVLOHIEYIESLEKAGILEEKMHLHDAVQDGLKRLVRNPPTVKIKIGELISMHPFLGALPQSVGRDLVQSTKEEVK	708
SVB9-10_S051_C6A0000001-665	625	SACYICAALRAHRTARGLHDFIOYSEISSAVITESTETEGEARKFLEDRVTITPEVLRVKTRQVTVAYVLOHIEYIESLEKAGILEEKMHLHDAVQDGLKRLVRNPPTVKIKIGELISMHPFLGALPQSVGRDLVQSTKEEVK	695
Quinoa_S051_C6A1-1281	898	ROMTLKYEOKPQIWLINSQVUKWACKVKKHKLHQTIVTGSTLQLYEVLQKPYLCMDITDSVAVGYIETEKILAALSDPAVENHFWKESVILAKVLLPRUFENUSMODMRKLTARSTLNTYLRBETIEVSSHIGFLLEQFV	1047
Quinoa_S051_C6A1-1211	829	ROMTLKYEOKPQIWLINSQVUKWACKVKKHKLHQTIVTGSTLQLYEVLQKPYLCMDITDSVAVGYIETEKILAALSDPAVENHFWKESVILAKVLLPRUFENUSMODMRKLTARSTLNTYLRBETIEVSSHIGFLLEQFV	977
Chenut_S051_C6A1-1152	769	ROMTLKYEOKPQIWLINSQVUKWACKVKKHKLHQTIVTGSTLQLYEVLQKPYLCMDITDSVAVGYIETEKILAALSDPAVENHFWKESVILAKVLLPRUFENUSMODMRKLTARSTLNTYLRBETIEVSSHIGFLLEQFV	918
SVB9-10_S051_C6A0000001-665	769	ROMTLKYEOKPQIWLINSQVUKWACKVKKHKLHQTIVTGSTLQLYEVLQKPYLCMDITDSVAVGYIETEKILAALSDPAVENHFWKESVILAKVLLPRUFENUSMODMRKLTARSTLNTYLRBETIEVSSHIGFLLEQFV	918
Quinoa_S051_C6A1-1281	1048	KSHPLAEELIPSPAALWPAQGNSSFLSQEESQYKSTSF.LHQOTSYYVETARVLLIDMVPICADNTLLRRKSSLLLDQSSRSLSSRDHAGLLSWPENQVKSQHLPDQGEIDDSQNLKSAKAMRLSIYGSTAKQVPVRGLSFQGYTLQNR	1197
Quinoa_S051_C6A1-1211	978	KSHPLAEELIPSPAALWPAQGNSSFLSQEESQYKSTSF.LHQOTSYYVETARVLLIDMVPICADNTLLRRKSSLLLDQSSRSLSSRDHAGLLSWPENQVKSQHLPDQGEIDDSQNLKSAKAMRLSIYGSTAKQVPVRGLSFQGYTLQNR	1127
Chenut_S051_C6A1-1152	919	KSHPLAEELIPSPAALWPAQGNSSFLSQEESQYKSTSF.LHQOTSYYVETARVLLIDMVPICADNTLLRRKSSLLLDQSSRSLSSRDHAGLLSWPENQVKSQHLPDQGEIDDSQNLKSAKAMRLSIYGSTAKQVPVRGLSFQGYTLQNR	1068
SVB9-10_S051_C6A0000001-665	919	KSHPLAEELIPSPAALWPAQGNSSFLSQEESQYKSTSF.LHQOTSYYVETARVLLIDMVPICADNTLLRRKSSLLLDQSSRSLSSRDHAGLLSWPENQVKSQHLPDQGEIDDSQNLKSAKAMRLSIYGSTAKQVPVRGLSFQGYTLQNR	1068
Quinoa_S051_C6A1-1281	1198	SHVRSYQGPVIGOKQLPLTSVKSEGSNTVRKRLQEDVMREELLPTTHSRHPSRAVDSSSESQEDVEFVRIDSPSKLSFRQAF	1281
Quinoa_S051_C6A1-1211	1128	SHVRSYQGPVIGOKQLPLTSVKSEGSNTVRKRLQEDVMREELLPTTHSRHPSRAVDSSSESQEDVEFVRIDSPSKLSFRQAF	1211
Chenut_S051_C6A1-1152	1059	SHVRSYQGPVIGOKQLPLTSVKSEGSNTVRKRLQEDVMREELLPTTHSRHPSRAVDSSSESQEDVEFVRIDSPSKLSFRQAF	1152
SVB9-10_S051_C6A0000001-665	1059	SHVRSYQGPVIGOKQLPLTSVKSEGSNTVRKRLQEDVMREELLPTTHSRHPSRAVDSSSESQEDVEFVRIDSPSKLSFRQAF	1152

B.



C.

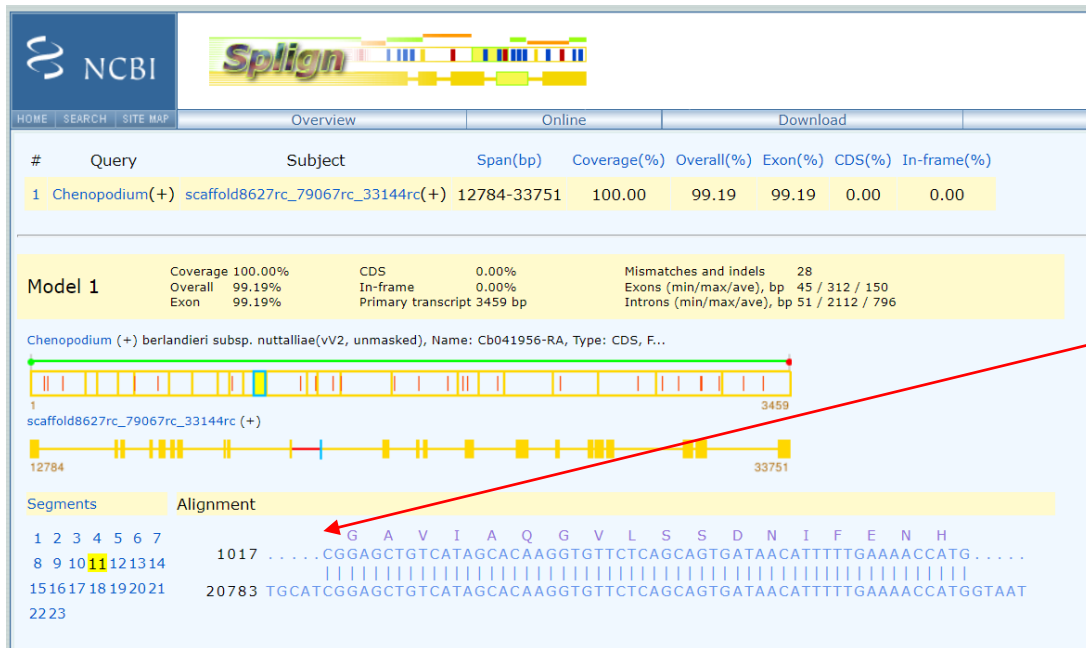


Figure S10. Gene models of *SOS1* sub-genome A. In all panels, the sequence whose name begins SV89-10 is from the *de novo* assembly reported here. A. Multiple sequence alignment of gene models from CoGe database for *C. quinoa* sub-genome A, *C. berlandieri nuttalliae*, and *C. pallidicaule* (the sub-genome 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.

A.

<i>Quinoa_CQ029119_C5A/1-362</i>	1MDNYLVGNKIGKNIRKSLPHQPNFVNSGTNNNGNNNTARQGPQPVYNI	96
<i>Capillidacule_CQ00195/1-372</i>	1	MDNSNMSKNRPNDYLVGNKIGKNIRKSLPHQPNFVNSGTNNNGNNNTARQGPQPVYNI	100
<i>Chefandienutalliae_Cb039629_C5A/1-372</i>	1	MDNSNMSKNRPNDYLVGNKIGKNIRKSLPHQPNFVNSGTNNNGNNNTARQGPQPVYNI	100
<i>Scaffold 7576/1-372</i>	1	MDNSNMSKNRPNDYLVGNKIGKNIRKSLPHQPNFVNSGTNNNGNNNTARQGPQPVYNI	100
<i>Quinoa_CQ029119_C5A/1-362</i>	97	INRRPMPVVAAPAPMQMPPFGQGHMAAPQVFNRRPPFGQGRSRMGQFAPGDAAWNPA	202
<i>Capillidacule_CQ00195/1-372</i>	107	INRRPMPVVAAPAPMQMPPFGQGHMAAPQVFNRRPPFGQGRSRMGQFAPGDAAWNPA	212
<i>Chefandienutalliae_Cb039629_C5A/1-372</i>	107	INRRPMPVVAAPAPMQMPPFGQGHMAAPQVFNRRPPFGQGRSRMGQFAPGDAAWNPA	212
<i>Scaffold 7576/1-372</i>	107	INRRPMPVVAAPAPMQMPPFGQGHMAAPQVFNRRPPFGQGRSRMGQFAPGDAAWNPA	212
<i>Quinoa_CQ029119_C5A/1-362</i>	203	QDQAPGHIQSHQPMHPQHGPHGDAQGVHMPDLVPGHVAQHPPSGLLPNPNMPPG	308
<i>Capillidacule_CQ00195/1-372</i>	213	QDQAPGHIQSHQPMHPQHGPHGDAQGVHMPDLVPGHVAQHPPSGLLPNPNMPPG	318
<i>Chefandienutalliae_Cb039629_C5A/1-372</i>	213	QDQAPGHIQSHQPMHPQHGPHGDAQGVHMPDLVPGHVAQHPPSGLLPNPNMPPG	318
<i>Scaffold 7576/1-372</i>	213	QDQAPGHIQSHQPMHPQHGPHGDAQGVHMPDLVPGHVAQHPPSGLLPNPNMPPG	318
<i>Quinoa_CQ029119_C5A/1-362</i>	300	FPPLLSNFAFSFQGSQGLGPGQPPLSPGYAFPLSPSGLF	362
<i>Capillidacule_CQ00195/1-372</i>	310	FPPLLSNFAFSFQGSQGLGPGQPPLSPGYAFPLSPSGLF	372
<i>Chefandienutalliae_Cb039629_C5A/1-372</i>	310	FPPLLSNFAFSFQGSQGLGPGQPPLSPGYAFPLSPSGLF	372
<i>Scaffold 7576/1-372</i>	310	FPPLLSNFAFSFQGSQGLGPGQPPLSPGYAFPLSPSGLF	372

B.

<i>Scaffold 1300/1-377</i>	1	MDNSNMSKNRPNDYLVGNKIGKNIRKSLPHQPNFVNSGTNNNGNNNTARQGPQPVYNI	96
<i>Capillidacule_CQ00195/1-372</i>	1	MDNSNMSKNRPNDYLVGNKIGKNIRKSLPHQPNFVNSGTNNNGNNNTARQGPQPVYNI	100
<i>Quinoa_CQ003751_C5B/1-374</i>	1	MDNSNMSKNRPNDYLVGNKIGKNIRKSLPHQPNFVNSGTNNNGNNNTARQGPQPVYNI	100
<i>Chefandienutalliae_Cb007314_C5B/1-375</i>	1	MDNSNMSKNRPNDYLVGNKIGKNIRKSLPHQPNFVNSGTNNNGNNNTARQGPQPVYNI	100
<i>Scaffold 1300/1-377</i>	152	POFGQGRSRMGQFAPGDAAWNPA	302
<i>Capillidacule_CQ00195/1-372</i>	140	POFGQGRSRMGQFAPGDAAWNPA	296
<i>Quinoa_CQ003751_C5B/1-374</i>	140	POFGQGRSRMGQFAPGDAAWNPA	296
<i>Chefandienutalliae_Cb007314_C5B/1-375</i>	150	POFGQGRSRMGQFAPGDAAWNPA	302
<i>Scaffold 1300/1-377</i>	303	FFNMSPTPRSPVPLSPGSGFPPLSPNFAFSFQGSQGLGPGQPPLSPGYAF	377
<i>Capillidacule_CQ00195/1-372</i>	300	FFNMSPTPRSPVPLSPGSGFPPLSPNFAFSFQGSQGLGPGQPPLSPGYAF	374
<i>Quinoa_CQ003751_C5B/1-374</i>	300	FFNMSPTPRSPVPLSPGSGFPPLSPNFAFSFQGSQGLGPGQPPLSPGYAF	374
<i>Chefandienutalliae_Cb007314_C5B/1-375</i>	301	FFNMSPTPRSPVPLSPGSGFPPLSPNFAFSFQGSQGLGPGQPPLSPGYAF	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.

<i>IKU2_subgenomeA/1-924</i>	1	MRPEPPTPLLLFLVLVCSNSASSDELNTLLTKSSLOSNPKFLSTWIPSNPTQFTVGSCHTARN	100
<i>Quinoa_CQ023322_C18/1-924</i>	1	MRPEPPTPLLLFLVLVCSNSASSDELNTLLTKSSLOSNPKFLSTWIPSNPTQFTVGSCHTARN	100
<i>Capillidacule_BBB1274/Intended/Spruce_paludicapsiae/1-924</i>	1	MRPEPPTPLLLFLVLVCSNSASSDELNTLLTKSSLOSNPKFLSTWIPSNPTQFTVGSCHTARN	100
<i>Chefandienutalliae_Cb039629_C5A/1-372</i>	1	MRPEPPTPLLLFLVLVCSNSASSDELNTLLTKSSLOSNPKFLSTWIPSNPTQFTVGSCHTARN	100
<i>Quinoa_CQ023322_C18/1-924</i>	151	DFSGKFWNLSNMNLTGILNSLODPFDESTFPQVTEKLNKLLNLTNTSSISGSIPEI	301
<i>Capillidacule_Cb039629_C5A/1-372</i>	151	DFSGKFWNLSNMNLTGILNSLODPFDESTFPQVTEKLNKLLNLTNTSSISGSIPEI	301
<i>Quinoa_CQ023322_C18/1-924</i>	151	DFSGKFWNLSNMNLTGILNSLODPFDESTFPQVTEKLNKLLNLTNTSSISGSIPEI	301
<i>Chefandienutalliae_Cb039629_C5A/1-372</i>	151	DFSGKFWNLSNMNLTGILNSLODPFDESTFPQVTEKLNKLLNLTNTSSISGSIPEI	301
<i>IKU2_subgenomeA/1-924</i>	302	PAEFQDFKLVNLSLVSNLTGOLPKOLSWTEFDFIDVSTNLSGPIPPDMCKKQTKMLL	402
<i>Quinoa_CQ023322_C18/1-924</i>	302	PAEFQDFKLVNLSLVSNLTGOLPKOLSWTEFDFIDVSTNLSGPIPPDMCKKQTKMLL	402
<i>Capillidacule_BBB1274/Intended/Spruce_paludicapsiae/1-924</i>	302	PAEFQDFKLVNLSLVSNLTGOLPKOLSWTEFDFIDVSTNLSGPIPPDMCKKQTKMLL	402
<i>Chefandienutalliae_Cb039629_C5A/1-372</i>	302	PAEFQDFKLVNLSLVSNLTGOLPKOLSWTEFDFIDVSTNLSGPIPPDMCKKQTKMLL	402
<i>IKU2_subgenomeA/1-924</i>	403	SSLLLDLNNKLVGEPESISIKMLKNNLVKQNFQSNIPVLSBECHLTDVNLGANSFSR	503
<i>Quinoa_CQ023322_C18/1-924</i>	403	SSLLLDLNNKLVGEPESISIKMLKNNLVKQNFQSNIPVLSBECHLTDVNLGANSFSR	503
<i>Capillidacule_BBB1274/Intended/Spruce_paludicapsiae/1-924</i>	403	SSLLLDLNNKLVGEPESISIKMLKNNLVKQNFQSNIPVLSBECHLTDVNLGANSFSR	503
<i>Chefandienutalliae_Cb039629_C5A/1-372</i>	403	SSLLLDLNNKLVGEPESISIKMLKNNLVKQNFQSNIPVLSBECHLTDVNLGANSFSR	503
<i>IKU2_subgenomeA/1-924</i>	604	FVILVLGDLALVALSGCYLTKLSGVGDKERSFLKDDYWNMSXVHLLTTEEL	704
<i>Quinoa_CQ023322_C18/1-924</i>	604	FVILVLGDLALVALSGCYLTKLSGVGDKERSFLKDDYWNMSXVHLLTTEEL	704
<i>Capillidacule_BBB1274/Intended/Spruce_paludicapsiae/1-924</i>	604	FVILVLGDLALVALSGCYLTKLSGVGDKERSFLKDDYWNMSXVHLLTTEEL	704
<i>Chefandienutalliae_Cb039629_C5A/1-372</i>	604	FVILVLGDLALVALSGCYLTKLSGVGDKERSFLKDDYWNMSXVHLLTTEEL	704
<i>IKU2_subgenomeA/1-924</i>	705	SLLVVEYMPNGLWDLRHSNNKLLADWDSRYEIALGAAGKLEYLHGVQKVPV	805
<i>Quinoa_CQ023322_C18/1-924</i>	705	SLLVVEYMPNGLWDLRHSNNKLLADWDSRYEIALGAAGKLEYLHGVQKVPV	805
<i>Capillidacule_BBB1274/Intended/Spruce_paludicapsiae/1-924</i>	705	SLLVVEYMPNGLWDLRHSNNKLLADWDSRYEIALGAAGKLEYLHGVQKVPV	805
<i>Chefandienutalliae_Cb039629_C5A/1-372</i>	705	SLLVVEYMPNGLWDLRHSNNKLLADWDSRYEIALGAAGKLEYLHGVQKVPV	805
<i>IKU2_subgenomeA/1-924</i>	800	SIIVGQDAKKIEESKEGF	904
<i>Quinoa_CQ023322_C18/1-924</i>	800	SIIVGQDAKKIEESKEGF	904
<i>Capillidacule_BBB1274/Intended/Spruce_paludicapsiae/1-924</i>	800	SIIVGQDAKKIEESKEGF	904
<i>Chefandienutalliae_Cb039629_C5A/1-372</i>	800	SIIVGQDAKKIEESKEGF	904

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.

Samuels Genomic sequence of Canadian *Chenopodium berlandieri*

A.

[illegible]

B.

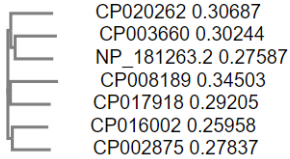
[illegible]

C.

Phylogenetic Tree

This is a Neighbour-joining tree without distance corrections.

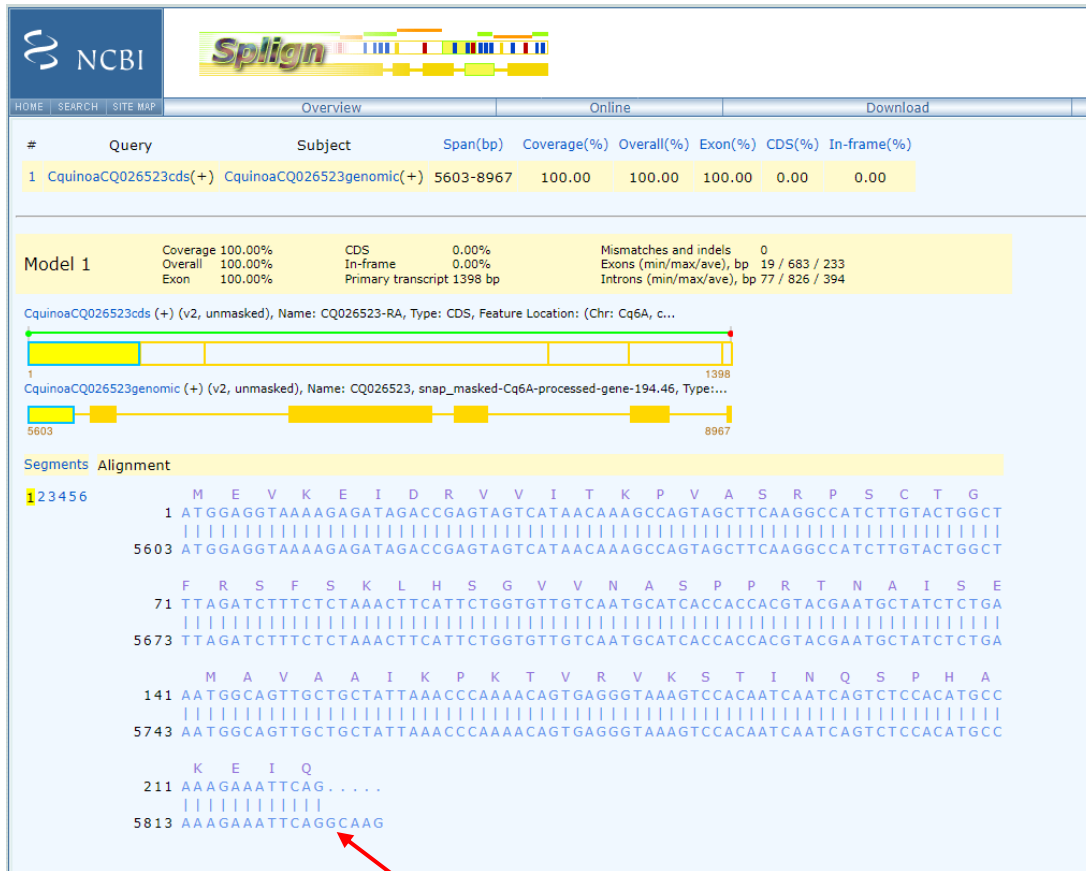
Branch length: ☐ Cladogram ☒ Real



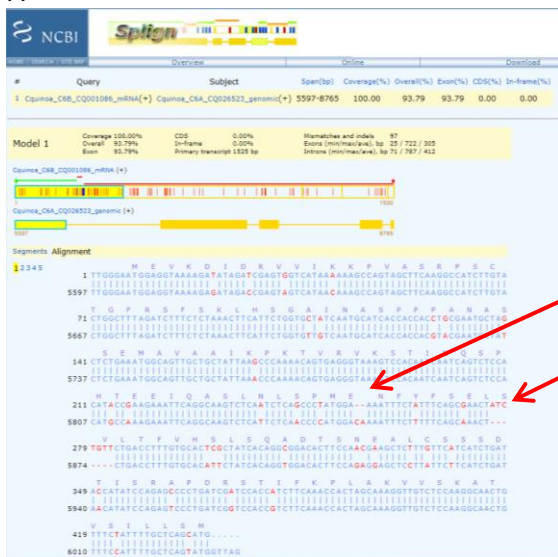
D.

[illegible]

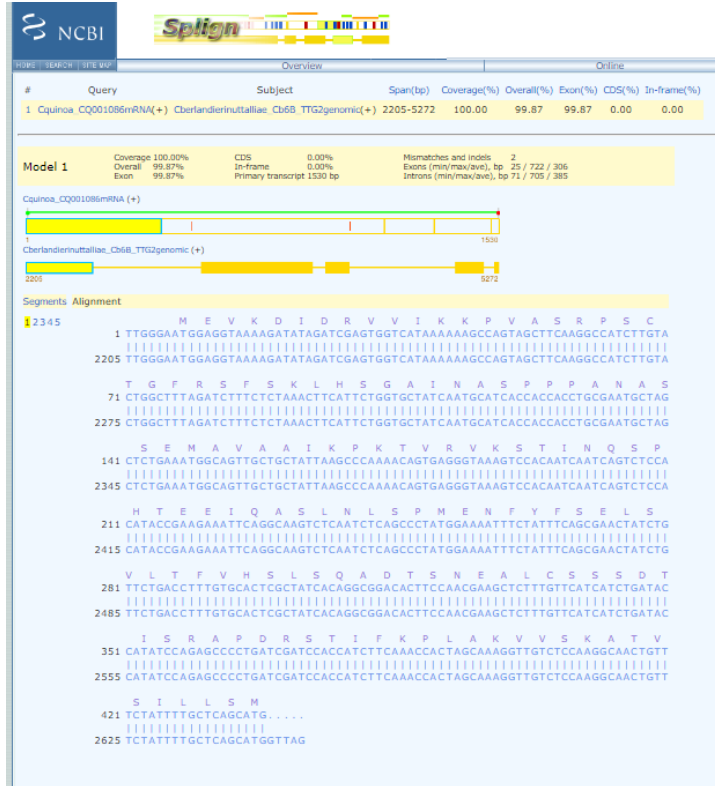
E.



F.



G.



H.

Gpallidicaule_CP003860/1-465 1 MEVKEIDRVVITKPVASRPSCTOFRSFSLKLSGVNNAASPPANAISEMAVAAIKPKTVRVKSTINQSPHAKETQAN.....TSGAPCSDSDNISRVPRSTVFK 100
 CuscutaBBB19587/1-499 1 MEVKEIDRVVITKPVASRPSCTOFRSFSLKLSGVNNAASPPANAISEMAVAAIKPKTVRVKSTINQSPHAKETQAN.....TSGAPCSDSDNISRVPRSTVFK 126
 Cberlandier_caffold27337/1-507 1 MEVKDIDRVVITKPVASRPSCTOFRSFSLKLSGVNNAASPPANAISEMAVAAIKPKTVRVKSTINQSPHAKETQAN.....TSGAPCSDSDNISRVPRSTVFK 127
 Quinoa_CQ001086_C6B/1-507 1 MEVKDIDRVVITKPVASRPSCTOFRSFSLKLSGVNNAASPPANAISEMAVAAIKPKTVRVKSTINQSPHAKETQAN.....TSGAPCSDSDNISRVPRSTVFK 127
 Cberlandierinuttalliae_Cb002931_C6B/1-507 1 MEVKDIDRVVITKPVASRPSCTOFRSFSLKLSGVNNAASPPANAISEMAVAAIKPKTVRVKSTINQSPHAKETQAN.....TSGAPCSDSDNISRVPRSTVFK 127

Gpallidicaule_CP003860/1-465 101 PLAKVVSATVSIILLSMONSSVTHROALSOVNVPAQKEKLHLFSSOHSSNLPVELAESAKSPTMAQQMVEGESRIMASSITGADQPSDQGNWRKYQKQVKQSGYPRSYKCTHLKCPVKKKVER 214
 CuscutaBBB19587/1-499 127 PLAKVVSATVSIILLSMONSSVTHROALSOVNVPAQKEKLHLFSSOHSSNLPVELAESAKSPTMAQQMVEGESRIMASSITGADQPSDQGNWRKYQKQVKQSGYPRSYKCTHLKCPVKKKVER 253
 Cberlandier_caffold27337/1-507 128 PLAKVVSATVSIILLSMONSSVTHROALSOVNVPAQKEKLHLFSSOHSSNLPVELAESAKSPTMAQQMVEGESRIMASSITGADQPSDQGNWRKYQKQVKQSGYPRSYKCTHLKCPVKKKVER 254
 Quinoa_CQ001086_C6B/1-507 128 PLAKVVSATVSIILLSMONSSVTHROALSOVNVPAQKEKLHLFSSOHSSNLPVELAESAKSPTMAQQMVEGESRIMASSITGADQPSDQGNWRKYQKQVKQSGYPRSYKCTHLKCPVKKKVER 254
 Cberlandierinuttalliae_Cb002931_C6B/1-507 128 PLAKVVSATVSIILLSMONSSVTHROALSOVNVPAQKEKLHLFSSOHSSNLPVELAESAKSPTMAQQMVEGESRIMASSITGADQPSDQGNWRKYQKQVKQSGYPRSYKCTHLKCPVKKKVER 254

Gpallidicaule_CP003860/1-465 215 SLDDQITEIVYKGEHNNPKPDHAKRSSSGVHGQELFAGTNQDPNIDPKLHDNVCEKNEOFERQMSNENEMQLSKOSTFPVNDAPVCHPVDAAEKHNPSDKTVVDSAGLGGGCEESAKGLNANOGRIR 341
 CuscutaBBB19587/1-499 254 SLDDQITEIVYKGEHNNPKPDHAKRSSSGVHGQELFAGTNQDPNIDPKLHDNVCEKNEOFERQMSNENEMQLSKOSTFPVNDAPVCHPVDAAEKHNPSDKTVVDSAGLGGGCEESAKGLNANOGRIR 380
 Cberlandier_caffold27337/1-507 255 SLDDQITEIVYKGEHNNPKPDHAKRSSSGVHGQELFAGTNQDPNIDPKLHDNVCEKNEOFERQMSNENEMQLSKOSTFPVNDAPVCHPVDAAEKHNPSDKTVVDSAGLGGGCEESAKGLNANOGRIR 381
 Quinoa_CQ001086_C6B/1-507 255 SLDDQITEIVYKGEHNNPKPDHAKRSSSGVHGQELFAGTNQDPNIDPKLHDNVCEKNEOFERQMSNENEMQLSKOSTFPVNDAPVCHPVDAAEKHNPSDKTVVDSAGLGGGCEESAKGLNANOGRIR 381
 Cberlandierinuttalliae_Cb002931_C6B/1-507 255 SLDDQITEIVYKGEHNNPKPDHAKRSSSGVHGQELFAGTNQDPNIDPKLHDNVCEKNEOFERQMSNENEMQLSKOSTFPVNDAPVCHPVDAAEKHNPSDKTVVDSAGLGGGCEESAKGLNANOGRIR 381

Gpallidicaule_CP003860/1-465 342 NKRKKQNVQNETGKSGEGVQEPGIWVHNINEPEIMDDGFRWRKYGQKVVKGSPYPRSYRCTSVKCNVRKYVERALDDSRIFTTYEGRHNHEMPMKIMNLAAVPEGETOPPAKEIRTLMPIVS 485
 CuscutaBBB19587/1-499 381 NKRKKQNVQNETGKSGEGVQEPGIWVHNINEPEIMDDGFRWRKYGQKVVKGSPYPRSYRCTSVKCNVRKYVERALDDSRIFTTYEGRHNHEMPMKIMNLAAVPEGETOPPAKEIRTLMPIVS 489
 Cberlandier_caffold27337/1-507 382 NKRKKQNVQNETGKSGEGVQEPGIWVHNINEPEIMDDGFRWRKYGQKVVKGSPYPRSYRCTSVKCNVRKYVERALDDSRIFTTYEGRHNHEMPMKIMNLAAVPEGETOPPAKEIRTLMPIVS 507
 Quinoa_CQ001086_C6B/1-507 382 NKRKKQNVQNETGKSGEGVQEPGIWVHNINEPEIMDDGFRWRKYGQKVVKGSPYPRSYRCTSVKCNVRKYVERALDDSRIFTTYEGRHNHEMPMKIMNLAAVPEGETOPPAKEIRTLMPIVS 507
 Cberlandierinuttalliae_Cb002931_C6B/1-507 382 NKRKKQNVQNETGKSGEGVQEPGIWVHNINEPEIMDDGFRWRKYGQKVVKGSPYPRSYRCTSVKCNVRKYVERALDDSRIFTTYEGRHNHEMPMKIMNLAAVPEGETOPPAKEIRTLMPIVS 507

I.

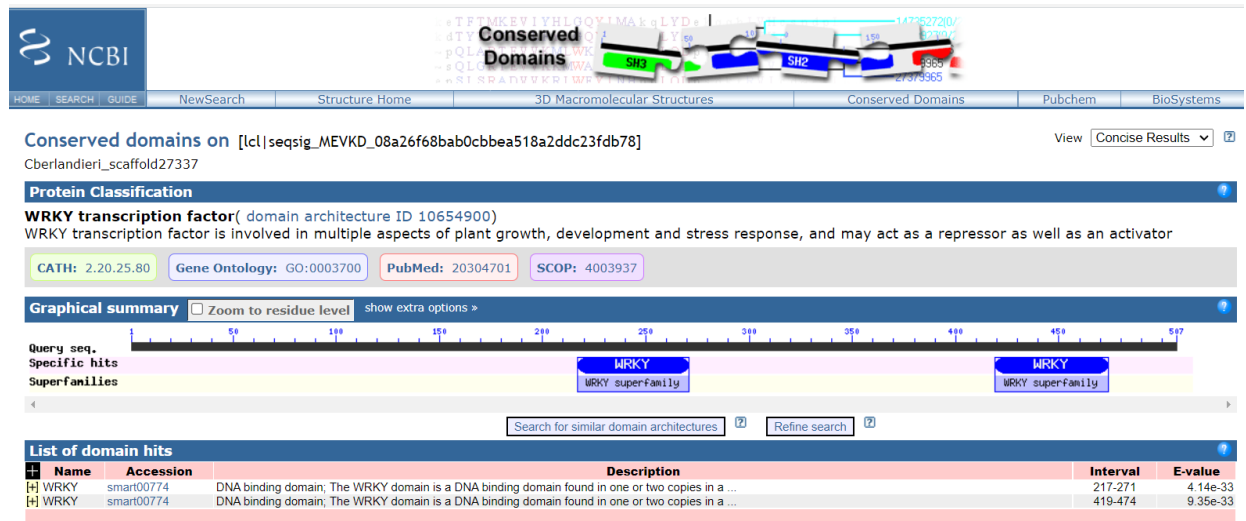


Figure S13. *TTG2* gene models. A. Potential homologs of *A. thaliana TTG2* (NP_181263) in *C. suecicum*. B. Potential homologs of *A. thaliana TTG2* (NP_181263) in *C. pallidicaule*. C. Clustal Phylogeny derived tree for potential homologs of *TTG2* in *C. pallidicaule*. D. C-terminus of *C. suecicum* 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 nuttalliae* 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 nuttalliae*, 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.