BMC Bioinformatics

BioMed Central

Software

Open Access

DNAAlignEditor: DNA alignment editor tool

Hector Sanchez-Villeda*[1], Steven Schroeder[1,4], Sherry Flint-Garcia[2], Katherine E Guill[2], Masanori Yamasaki[1,3] and Michael D McMullen[1,2]

Address: [1]Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA, [2]USDA-ARS, Plant Genetics Research Unit, Columbia, MO 65211, USA, [3]Food Resources Education and Research Center, Kobe University, Hyogo, Japan and [4]USDA-ARS Bovine Functional Genomics Laboratory, Beltsville, MD 20705, USA

Email: Hector Sanchez-Villeda* - sanchezvilledah@missouri.edu; Steven Schroeder - steven.schroeder@ars.usda.gov; Sherry Flint-Garcia - Sherry.Flint-Garcia@ars.usda.gov; Katherine E Guill - Kate.Guill@ars.usda.gov; Masanori Yamasaki - yamasakim@tiger.kobe-u.ac.jp; Michael D McMullen - Michael.McMullen@ars.usda.gov

* Corresponding author

Published: 19 March 2008

BMC Bioinformatics 2008, **9**:154 doi:10.1186/1471-2105-9-154

Received: 4 October 2007
Accepted: 19 March 2008

This article is available from: http://www.biomedcentral.com/1471-2105/9/154

Abstract

Background: With advances in DNA re-sequencing methods and Next-Generation parallel sequencing approaches, there has been a large increase in genomic efforts to define and analyze the sequence variability present among individuals within a species. For very polymorphic species such as maize, this has lead to a need for intuitive, user-friendly software that aids the biologist, often with naïve programming capability, in tracking, editing, displaying, and exporting multiple individual sequence alignments. To fill this need we have developed a novel DNA alignment editor.

Results: We have generated a nucleotide sequence alignment editor (DNAAlignEditor) that provides an intuitive, user-friendly interface for manual editing of multiple sequence alignments with functions for input, editing, and output of sequence alignments. The color-coding of nucleotide identity and the display of associated quality score aids in the manual alignment editing process. DNAAlignEditor works as a client/server tool having two main components: a relational database that collects the processed alignments and a user interface connected to database through universal data access connectivity drivers. DNAAlignEditor can be used either as a stand-alone application or as a network application with multiple users concurrently connected.

Conclusion: We anticipate that this software will be of general interest to biologists and population genetics in editing DNA sequence alignments and analyzing natural sequence variation regardless of species, and will be particularly useful for manual alignment editing of sequences in species with high levels of polymorphism.

Background

Maize is a highly polymorphic species with numerous single nucleotide polymorphisms (SNP) and insertion/deletions (InDel) commonly present among maize alleles [1,2]. One consequence of this level of polymorphism is that sequence alignments of maize genes constructed with standard software such as CLUSTAL W, Muscle, Dialign, MAFFT, T-Coffee, etc., must be visually inspected and manually edited to ensure that an optimal alignment is obtained. This problem, together with the need to make base calls and alignment decisions in regions of variable base quality, creates the need for a tool to assist the researcher in manual editing of sequence alignments. Researchers in communities other than maize also require user-friendly tools to facilitate input, processing, and output of the sequence diversity among individuals within the species. Although there are several improved methods [3] to generate alignments for multiple nucleotide sequences, there is a need to improve software for manual editing to evaluate and fix discrepancies. Our goal was to produce a sequence alignment tool (DNAAlignEditor) that provides an efficient path for editing multiple sequence alignments to improve alignment quality, tracking edited alignments, and exporting sequence files to external tools for analysis.

Implementation

DNAAlignEditor is written in Visual Basic for the user interface, MS Access database for the stand-alone version or PostrgeSQL 8.0 database for the network version, and embedded ActiveX data object drivers for connectivity between the database and the user interface. DNAAlignEditor network version is designed to share alignment information when several users are simultaneously logged into the system and to manage alignments collected in the database. Regardless of whether the stand-alone or network version is used, the database aids in the tracking and manipulation of large numbers of sequences, and the relationship among genes, primers, amplicons, and population groups. The database is also required to track multiple versions of the same alignments and to combine alignments. Editing of a handful of gene alignments may be accomplished without the aid of a database; however, processing, editing, and finishing hundreds of alignments requires a more sophisticated system managed by a database.

DNAAlignEditor stores a set of interrelated tables in a relational database (DNADB) supporting alignment information to provide easy editing, querying, importing, and exporting function from genes, primers, populations, nucleotide sequence alignments, and quality scores as shown in the entity relationship diagram [see additional file 1]. The system is a user-friendly editing tool using grids, text boxes, combo boxes, and scroll lists to facilitate data access. DNAAlignEditor tool color codes nucleotide identity for ease in visual identification of sequence polymorphisms and displays the quality score associated with each base to aid in manual editing. The quality score-centric view is one of the major and unique features that distinguish DNAAlignEditor from other alignment programs (Figure 1).

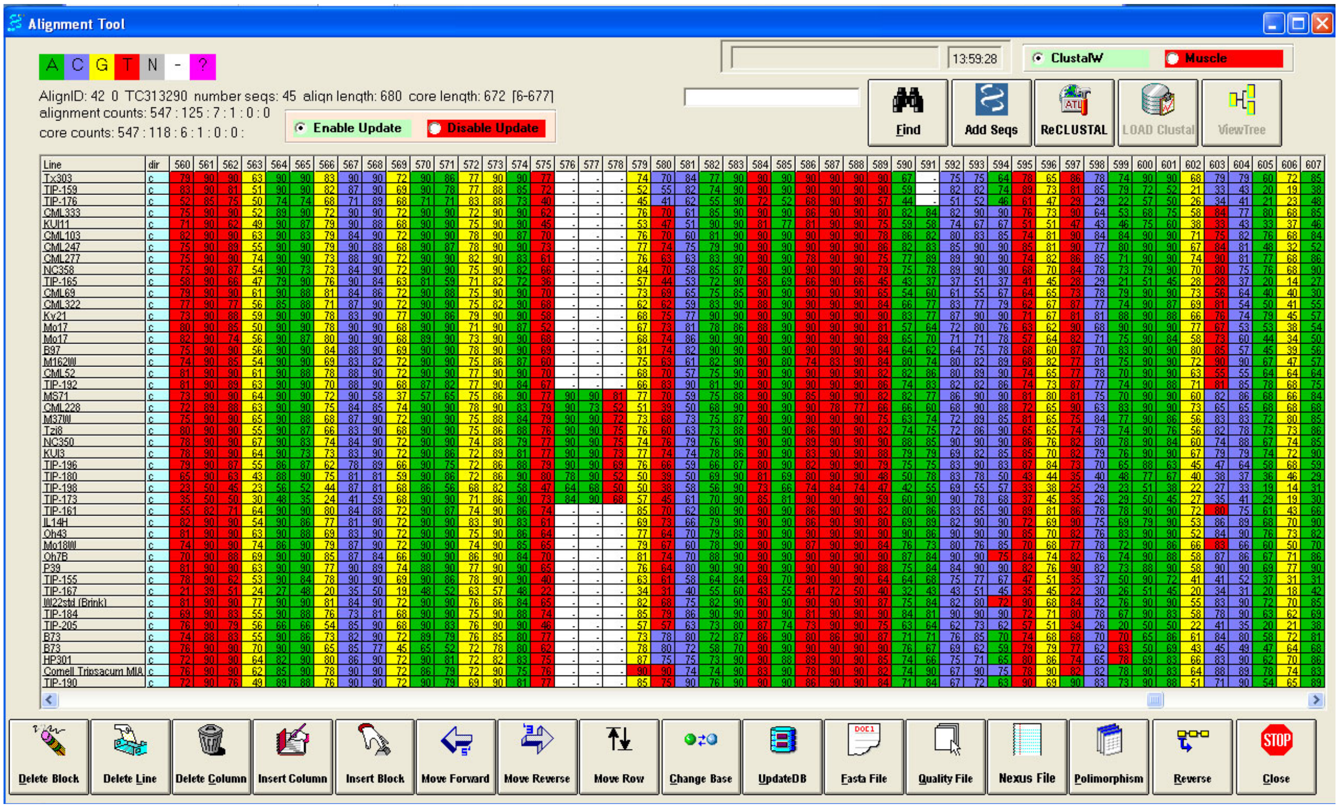
DNAAlignEditor accepts alignments in three different formats and checks the consistency of the information before addition to the database [see additional file 2]. The first format uses the information from PHRAP ace files, and a file that defines the gene information associated with the alignment. The second format is a fasta format. The last format accepts a tab delimited text, file having a header with information related to gene, amplicon, and total bases followed by a nucleotide sequence row and a quality scores row (or no-quality where a default quality score of 15 is applied to all base calls) [see additional file 2].

The program validates the data before it is imported, i.e. that genes, amplicons, and populations exist in DNADB. For sequences without quality scores, the program assigns a default quality score of 15 for each base. This value can be changed in the configuration file. By definition, a *phred* quality score of $Q \geq 20$ indicates that there is a less than 1 in 100 chance that the base call is incorrect. Consequently, "Q20" bases are considered "high quality" calls with corresponding low error probabilities [4,5]. By adding a standard quality score of $Q = 15$, we effectively flag those bases as unknown quality.

Once the alignment information is in the database, it can be edited in an intuitive, user-friendly interface. A user help file can be accessed by pressing the F1 key. The tool provides several features to edit information stored in the database such as moving individual sequences up and down for ease of fixing errors in multiple lines, moving bases from the 5' direction to the 3' direction and conversely, deleting and inserting individual bases and block of bases, and deleting and inserting entire columns in the alignment. DNAAlignEditor forms initial alignments by executing either of the integrated CLUSTAL W [6] or MUSCLE programs [7], and displays and prints phylogenies with TreeView [8]. The inclusion of a "Re-CLUSTAL" function allows editing to be preformed as a reiterative process, for instance rerunning CLUSTAL W after trimming end sequences or removing lower quality individual sequences for the initial alignment. DNAAlignEditor also allows the user to export fasta files with quality files and/ or files compatible with DnaSP [9] for further analysis. Upon editing changes and saving, the tool will generate a new alignment while maintaining the original for further use or reference.

Results and discussion

There are several software packages that produce nucleotide sequence alignments such as CLUSTAL W, Muscle, Dialign, MAFFT, and T-Coffee. However, all of these programs produce alignments that must be manually examined and edited for accuracy. By integrating a highly intuitive visual display with CLUSTAL functions, DNAAlignEditor allows the researcher move seamlessly between alignment and editing functions. The inclusion of quality scores allow the researcher to make informed decisions on editing InDel containing regions and accessing the certainty of unique nucleotide polymorphisms. The database



● ClustalW ○ Muscle

AlignID: 42 0 TC313290 number seqs: 45 align length: 680 core length: 672 [6-677]
alignment counts: 547 : 125 : 7 : 1 : 0 : 0
core counts: 547 : 118 : 6 : 1 : 0 : 0 :

● Enable Update ○ Disable Update

Line	dir	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607
Tx303	c	79	90	90	63	90	90	83	90	90	72	90	86	77	90	90	77	-	-	-	74	70	84	77	90	90	90	90	90	90	90	67	-	75	75	64	78	65	86	78	74	90	90	68	79	79	60	72	85
TIP-159	c	83	90	81	51	90	90	82	87	90	69	90	78	77	88	85	72	-	-	-	52	55	82	74	90	90	90	90	90	90	90	59	-	82	82	74	89	73	81	85	79	72	52	21	33	43	20	19	38
TIP-176	c	52	85	75	50	74	74	68	71	89	68	71	71	83	88	73	40	-	-	-	45	41	62	55	90	72	52	68	90	90	57	44	-	51	52	46	61	47	29	29	22	57	50	26	34	41	21	23	48
CML333	c	75	90	90	52	89	90	72	90	90	72	90	90	72	90	90	62	-	-	-	76	70	61	85	90	90	90	86	90	90	80	82	84	82	90	90	76	73	90	64	53	68	75	58	84	77	80	68	85
KUI11	c	71	90	62	49	90	87	79	90	88	68	90	90	75	90	90	45	-	-	-	53	47	51	90	90	81	77	81	90	90	75	59	58	74	67	67	51	51	47	43	46	75	60	38	33	43	33	37	46
CML103	c	82	90	90	63	90	83	79	84	90	72	90	90	78	90	87	70	-	-	-	76	70	60	81	90	90	90	90	90	90	78	86	82	80	83	85	74	81	90	84	84	90	90	71	75	82	76	68	84
CML247	c	75	90	89	55	90	90	79	90	88	68	90	87	78	90	90	73	-	-	-	77	74	75	79	90	90	90	90	90	90	86	82	83	85	90	90	85	81	90	77	80	90	90	67	84	81	48	32	52
CML277	c	75	90	90	74	90	90	73	88	90	72	90	90	82	90	83	61	-	-	-	76	63	63	83	90	90	90	78	90	90	75	77	89	89	90	90	74	82	86	85	71	90	90	74	90	81	77	68	86
NC358	c	75	90	87	54	90	73	73	84	90	72	90	90	75	90	82	66	-	-	-	84	70	58	85	87	90	90	90	90	90	79	75	78	89	90	90	68	70	84	78	73	79	90	70	80	75	76	68	90
TIP-165	c	58	90	66	47	79	90	76	90	84	63	81	59	71	82	72	36	-	-	-	57	44	53	72	90	58	69	66	90	66	45	43	37	37	51	37	41	45	28	29	21	51	45	28	28	37	20	14	27
CML69	c	79	90	90	61	90	88	81	84	86	72	90	88	75	90	90	70	-	-	-	73	69	65	75	85	90	90	90	90	90	65	54	60	61	55	67	64	65	73	78	79	90	90	73	56	64	40	40	30
CML322	c	77	90	77	56	85	88	71	87	90	72	90	90	75	82	90	68	-	-	-	62	62	59	83	90	88	90	90	90	90	84	66	77	83	77	79	62	67	87	77	74	90	87	69	81	54	50	41	55
Ky21	c	73	90	88	59	90	90	78	83	90	77	90	86	79	90	90	58	-	-	-	68	75	77	90	90	90	90	90	90	90	90	83	77	87	90	90	71	67	81	81	88	90	88	66	76	74	79	45	57
Mo17	c	80	90	85	50	90	90	78	90	90	68	90	90	71	90	87	52	-	-	-	67	73	81	78	86	88	90	90	90	90	81	57	64	72	80	76	63	62	90	68	90	90	90	77	67	53	53	38	54
Mo17	c	82	90	74	56	90	87	80	90	90	68	89	90	73	90	90	68	-	-	-	68	74	86	90	90	90	90	90	90	90	89	65	70	71	71	78	57	64	82	71	75	90	84	58	73	60	44	34	50
B97	c	75	90	90	56	90	90	84	88	90	69	90	90	78	90	90	69	-	-	-	81	74	82	90	90	90	90	90	90	90	84	64	62	64	75	78	68	60	87	70	83	90	90	80	85	57	45	39	56
M162W	c	74	90	85	54	90	90	69	83	82	72	90	90	75	86	87	60	-	-	-	75	63	61	82	90	90	80	74	83	90	84	80	74	80	82	89	68	82	77	81	75	90	90	72	90	90	67	47	57
CML52	c	81	90	90	61	90	88	78	88	90	72	90	90	77	90	90	70	-	-	-	68	70	57	75	90	90	90	90	90	90	82	82	86	80	89	90	74	65	77	78	70	90	90	63	55	55	64	64	64
TIP-192	c	81	90	89	63	90	90	70	88	90	68	87	82	77	90	84	67	-	-	-	66	83	90	81	90	90	90	90	90	90	86	74	83	82	82	86	74	73	87	77	74	90	88	71	81	85	78	68	75
MS71	c	73	90	90	64	90	90	72	90	58	37	57	65	75	86	90	77	90	90	81	77	70	59	75	88	90	90	85	90	90	82	82	77	86	90	90	81	80	81	75	70	90	90	60	82	86	68	66	84
CML228	c	72	89	88	63	90	90	75	84	85	74	90	90	78	90	83	79	90	73	52	51	39	50	68	90	90	90	90	78	77	66	66	60	68	90	88	72	65	90	63	83	90	90	73	65	65	68	68	68
M37W	c	75	90	90	65	90	88	68	87	90	72	90	90	75	88	84	79	90	90	72	73	68	73	75	87	90	90	90	90	90	75	63	74	72	89	85	81	65	75	84	77	90	86	56	83	83	72	80	85
Tzi8	c	80	90	90	55	90	87	66	83	90	68	90	90	75	86	88	76	90	90	75	76	60	63	73	88	90	90	86	90	90	82	74	75	72	86	90	65	65	74	73	74	90	76	56	82	78	73	73	86
NC350	c	78	90	90	67	90	83	74	84	90	72	90	90	74	88	79	77	90	90	75	74	76	79	76	90	90	90	89	90	90	90	88	85	90	90	90	86	76	82	80	78	90	84	60	74	88	67	74	85
KUI3	c	78	90	90	64	90	73	73	83	90	72	90	86	72	89	81	77	90	90	73	77	74	74	78	86	90	90	83	90	90	88	79	79	69	82	85	85	70	82	79	76	90	90	67	79	79	74	72	90
TIP-196	c	79	90	87	55	86	87	62	78	89	66	90	75	72	86	88	79	90	90	69	76	66	59	66	87	80	90	82	90	90	79	75	75	83	90	83	87	84	73	70	65	88	63	45	47	64	58	68	59
TIP-180	c	65	90	63	43	88	90	75	81	81	59	90	86	72	86	90	80	78	90	52	50	39	50	69	90	81	69	80	90	90	48	50	78	83	78	50	43	44	35	40	48	77	67	40	38	37	36	46	29
TIP-198	c	23	50	45	23	56	52	44	87	81	68	86	56	68	82	58	47	64	68	50	50	38	58	56	90	73	66	74	84	74	47	42	55	69	55	57	33	38	25	29	23	51	38	22	27	33	19	14	31
TIP-173	c	35	50	50	30	48	35	24	41	59	68	90	90	71	86	90	73	84	90	68	57	45	61	70	90	85	81	90	90	90	59	60	90	90	78	68	37	45	35	26	29	50	45	27	35	41	29	19	30
TIP-161	c	55	82	71	64	90	90	80	84	88	72	90	87	74	90	86	74	-	-	-	85	70	62	80	90	90	90	86	90	90	82	80	86	83	85	90	89	81	86	78	78	90	90	72	80	75	61	43	66
IL14H	c	82	90	90	54	90	86	77	81	90	72	90	90	83	90	83	61	-	-	-	69	73	66	79	90	90	90	86	90	90	80	69	89	82	90	90	72	69	90	75	69	79	90	53	86	89	68	70	90
Oh43	c	81	90	90	63	90	88	69	83	90	72	90	90	75	90	86	64	-	-	-	77	64	70	79	88	90	90	90	90	90	90	90	86	90	90	90	85	70	82	76	83	90	90	52	84	90	76	73	82
Mo18W	c	74	90	90	74	86	90	79	87	90	72	90	90	74	90	85	65	-	-	-	79	67	60	78	90	90	90	87	90	90	84	76	73	80	76	85	70	68	77	78	72	90	86	66	83	66	60	50	70
Oh7B	c	70	90	88	69	90	90	85	87	84	66	90	90	86	90	84	70	-	-	-	81	74	70	88	90	90	90	90	90	90	90	87	84	90	90	75	84	74	82	76	74	90	88	58	87	86	67	71	86
P39	c	81	90	90	63	90	90	77	90	89	74	88	90	77	90	90	65	-	-	-	76	64	80	90	90	90	90	90	90	90	88	75	84	84	90	90	82	76	90	82	73	88	90	58	90	90	69	77	90
TIP-155	c	78	90	62	53	90	84	78	90	90	69	90	86	78	90	90	40	-	-	-	63	61	58	64	84	69	70	90	90	90	64	64	68	75	77	67	47	51	35	37	50	90	72	41	41	52	37	31	31
TIP-167	c	21	39	51	24	27	48	20	35	50	19	48	52	63	57	48	22	-	-	-	34	31	40	55	60	43	55	41	72	50	40	32	43	43	51	45	35	45	22	30	26	51	45	20	34	31	20	18	42
W22std (Brink)	c	81	90	90	77	90	90	81	84	90	72	90	90	76	86	84	65	-	-	-	82	68	75	82	90	90	90	90	90	90	87	75	84	82	80	72	90	68	84	82	76	90	90	55	83	90	72	70	85
TIP-184	c	69	90	83	55	90	88	76	73	81	68	90	90	75	90	88	74	-	-	-	85	79	86	90	90	90	90	81	90	90	90	84	81	90	90	90	72	71	80	78	67	90	83	58	78	90	63	62	69
TIP-205	c	76	90	79	56	66	66	54	85	90	68	90	83	76	90	90	46	-	-	-	57	57	63	73	80	87	74	73	90	90	75	63	64	62	73	62	57	51	34	26	20	50	50	22	41	35	20	21	38
B73	c	74	88	83	55	90	86	73	82	90	72	89	79	76	85	80	77	-	-	-	73	78	80	72	87	86	90	80	86	90	87	71	71	76	85	70	74	68	68	70	70	65	86	61	84	80	58	72	81
B73	c	76	90	90	70	90	90	65	85	77	45	65	52	72	78	80	62	-	-	-	78	80	72	58	70	90	90	90	90	90	90	76	67	69	62	59	79	79	77	62	63	50	69	43	45	49	47	64	68
HP301	c	72	90	90	64	82	90	80	86	90	72	90	81	72	82	83	75	-	-	-	87	75	75	73	90	90	88	89	90	90	85	74	66	75	71	65	80	86	74	65	78	69	83	66	83	90	62	70	86
Cornell Tripsacum MIA	c	76	90	90	62	85	90	78	90	90	72	86	79	72	90	75	76	-	-	-	90	90	74	74	90	83	90	78	90	90	82	74	90	67	90	75	78	90	82	82	78	90	83	64	88	89	78	74	83
TIP-190	c	72	90	76	49	89	88	76	90	90	72	90	79	69	90	81	77	-	-	-	85	75	90	76	90	90	90	86	90	90	84	71	84	67	72	63	90	69	90	83	73	90	88	51	71	90	54	65	89

Figure 1

DNAAlignEditor view of an example alignment. The color coded cells represent the base calls and the numbers within the cells reflect the quality scores. Insertion/deletions are present in columns 576–578 and 591, and SNPs in columns 579, 580, 594, 597, 599 and 603. Below the alignment are the buttons used to manipulate sequence. To the left of the alignment are the names of the individuals displayed. Above the alignment are identification summary statistics on the alignment and additional buttons to manipulate sequence file and re-CLUSTAL functions.

functions promote efficient tracking of large sets of primers, genes, amplicons; and the relationships among these components of diversity sequencing projects. It is the focus on intuitive manual editing functions, inclusion of quality scores, and power of tracking components in large project that distinguish the software from other available for alignment functions. While a variety of alignment editors are now available include BioEdit [10], JalView [11], and INTERALIGN [12]. Of these only JalView, like DNAAlignEditor, is open source. None of the other programs is as intuitive to the naïve user nor uses the quality score feature central to DNAAlignEditor. In addition, the database functions of DNAAlignEditor for tracking the components of a re-sequencing projects makes DNAAlignEditor more appropriate for large DNA sequence diversity projects.

This software has been successfully used by the National Science Foundation funded "Molecular and Functional Diversity of the Maize Genome" group at the University of Missouri [13], University of Wisconsin, and Cold Spring Harbor Laboratories for manual editing of alignments. DNAAlignEditor was essential for tracking and exporting over 4,000 alignments are currently stored in DNADB with more than 50,000 sequences from several maize, teosinte, and Tripsacum accessions. In addition the amplicon merge feature was very helpful for extended sequencing of genes of particular interest to the group. The finished alignments are displayed in a comprehensive graphical user interface tool included in panzea site [14].

Conclusion

We have developed a nucleotide sequence alignment editor (DNAAlignEditor) that provides an intuitive, user-friendly tool for manual editing of multiple sequence alignments which can be used for different species to allow researchers to visualize and edit the alignment information in a straight forward way, and realign the sequences using the CLUSTAL W or Muscle software, as well as export information to different packages for further analysis.

Availability and requirements

Project name: DNAAlignEditor executable and source code are freely available at http://maize.agron.missouri.edu/~hsanchez/DNAAlignment_Tool.html [See additional files 3, 4]

Contact: sanchezvilledah@missouri.edu

Operating systems: Windows 95/98/NT/2000/XP; can run on Intel Mac using Parallels

Programming language: Visual Basic 6.0

Abbreviations

CLUSTAL W: general purpose multiple sequence alignment program for DNA or protein sequences. Muscle: multiple sequence alignment comparison by log-expectation program.

MAFFT: multiple sequence alignment program. T-COFFE: multiple sequence alignment program.

ActiveX: is the name Microsoft has given to a set of "strategic" object-oriented programming technologies and tools. SNP: single nucleotide polymorphism. InDel: insertion-deletion polymorphism. MS Access: Microsoft Access. PHRAP: is a program for assembling shotgun DNA sequence data. Phred: is a program that reads DNA sequencing trace files, calls bases, and assigns a quality value to each called base.

Authors' contributions

HS-V elaborated the design, performed programming, wrote the documentation for the SNPAlign Editor tool, and prepared the manuscript. SS assisted in programming and design. SF-G, KEG and MY contributed design ideas and testing. MDM contributed to design and supervised the project. All authors have read and approved the manuscript.

Additional material

Additional file 1

Entity-relationship diagram for DNADB database. The first column of each table contains the field name and the second column contains the data type. The underlined row indicates that the field is a primary key.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-9-154-S1.jpeg]

Additional file 2

DNA alignment editor tool file formats. Definition of file formats accepted to import DNA alignments to DNAAlignEditor tool.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-9-154-S2.doc]

Additional file 3

DNA alignment editor tool source core. Source code for DNA align editor tool in zip format.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-9-154-S3.zip]

Additional file 4

DNA alignment tool installation instructions. Installation steps for DNAAlignEditor tool.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-9-154-S4.doc]

Acknowledgements

We thank Allison Weber and Chris Maher for providing comments and interface suggestions, and testing the application. This research was supported by National Science Foundation Plant Genome Award DBI0321467, and funds provided to MDM from the USDA-ARS. The names of products are necessary to report factually on available data; however, neither the USDA nor the University of Missouri guarantees or warrants the standard of the product, and the use of the name does not imply the approval of the product to the exclusion of others that may also be suitable.

References

1. Vroh Bi I, McMullen MD, Sanchez-Villeda H, Schroeder S, Gardiner J, Polacco M, Soderlund C, Wing R, Fang Z, Coe EH Jr: **Single nucleotide polymorphisms and insertion-deletions for genetic markers and anchoring the maize fingerprint contig physical map.** *Crop Sci* 2006, **46:**12-21.
2. Wright SI, Vroh Bi I, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS: **The effects of artificial selection on the maize genome.** *Science* 2005, **308:**1310-1314.
3. Schwartz AS, Pachter L: **Multiple alignment by sequence annealing.** *Bioinformatics* 2006, **23(2):**e24-e29.
4. Ewing B, Green P: **Basecalling of automated sequencer traces using *phred*. II. Error probabilities.** *Genome Res* 1998, **8:**186-194.
5. Ewing B, Hillier L, Wendl M, Green P: **Basecalling of automated sequencer traces using phred. I. Accuracy assessment.** *Genome Res* 1998, **8:**175-185.
6. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucl Acids Res* 1994, **22:**4673-4680.
7. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucl Acids Res* 2004, **32:**1792-1797.
8. Page RDM: **TREEVIEW: An application to display phylogenetic trees on personal computers.** *Comp Appl Biosci* 1996, **12:**357-358.
9. Rozas J, Sánchez-DelBarrio SJ, Messeguer X, Rozas R: **DnaSP, DNA polymorphism analyses by the coalescent and other methods.** *Bioinformatics* 2003, **19:**2496-2497.
10. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucl Acids Symp Ser* 1999, **41:**95-98.
11. Clamp M, Cuff J, Searle SM, Barton JG: **The Jalview Java alignment editor.** *Bioinformatics* 2004, **20:**426-427.
12. Pible O, Imbert G, Pellequer J-L: **INTERALIGN: interactive alignment editor for distantly related protein sequences.** *Bioinformatics* 2005, **21:**3166-3167.
13. Yamasaki M, Tenaillon MI, Vroh Bi I, Schroeder SG, Sanchez-Villeda H, Doebley JF, Gaut BS, McMullen MD: **A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement.** *Plant Cell* 2005, **17:**2859-2872.

14. Canaran P, Stein L, Ware D: **Look-Align: An interactive web-based multiple sequence alignment viewer with polymorphism analysis support.** *Bioinformatics* 2006, **22:**885-886.