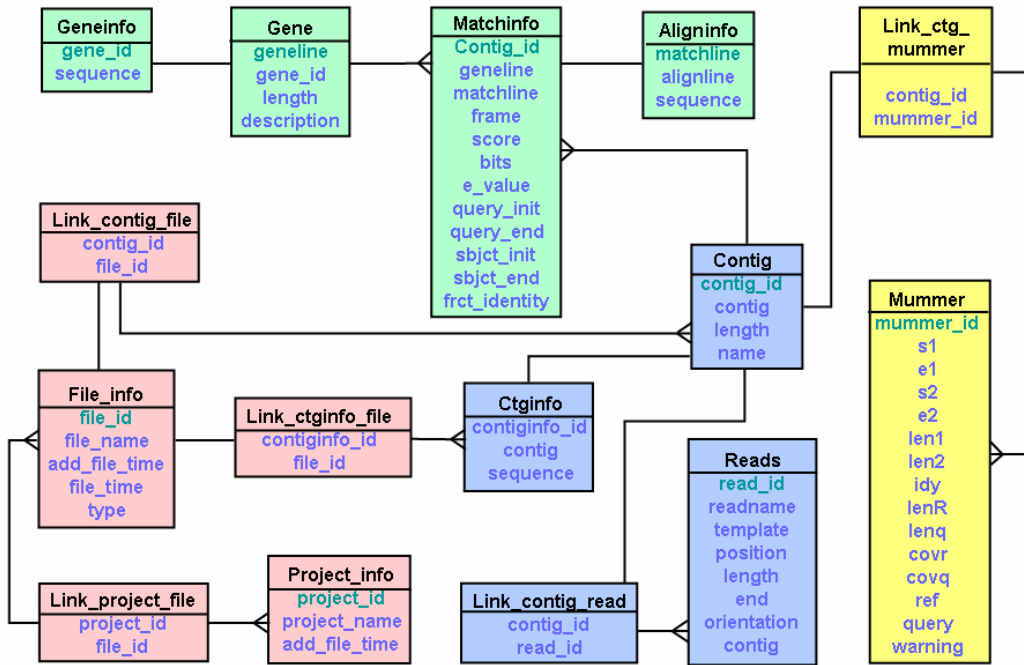


1 **Supplementary data**

2



3

4 **Fig: S1:** The entity-relationship model for TOPAASdb. MySQL table schemes are
 5 represented by color coded table diagrams and describe the Project (red), Blast (green),
 6 MUMmer (yellow), and Assembly (blue) entity class. The name of each table is given in bold,
 7 key attributes are given in green colored font, and normal attributes are given below in blue
 8 colored font. The relationship between entity classes is indicated by connectors. The
 9 Link_ctg_mummer, Link_ctg_read, Link_ctginfo_file, and Link_project_file table
 10 schemes are included to visualize the breakdown of the relationships between entity classes.

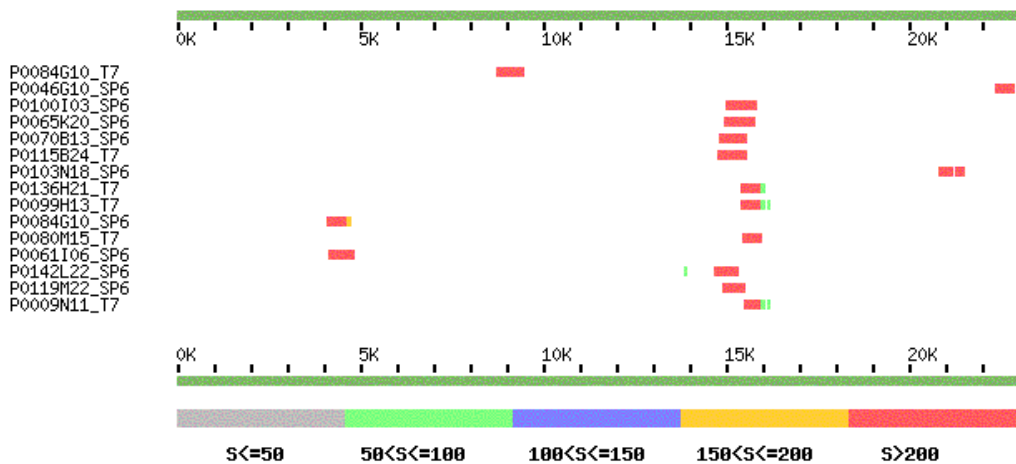
11

Primers	General Conditions	Other Conditions	Penalty Weights
Primers			
contig	side	primer sequence	Tm blast results start length
<input checked="" type="checkbox"/> RH123XF00380 (19436 bp)	left side	TGGGAAGAGTTGCAATTGTG	59.9 1 120 20
	right side	TCGAGCATCTTCAAGTGGTTG	60.0 1 19005 20
<input checked="" type="checkbox"/> RH123XR00149 (32847 bp)	left side	TGCCACATGGACACATTAC	60.2 1 103 20
	right side	GCGGTATTGTTTATGTCATGG	59.0 1 32311 21
<input checked="" type="checkbox"/> RH123XR00875 (21108 bp)	left side	TCATACATTGCGGAAAATG	59.4 1 247 20
	right side	TTTGGTCGGATTGTGTTG	59.4 2 20547 20
<input checked="" type="checkbox"/> RH123XR01119 (57034 bp)	left side	CAGGAAAATCATCTTGGTTGG	59.4 1 201 21
	right side	CCCTCATGTTCTGTTCAAG	60.7 1 56634 20
Generate		1 primer(s) for checked contigs	
		Export to file	

1
2 **Fig. S2:** Snapshot of a TOPAAS primer table for potato BAC contigs from RH123P09.
3 Colored connecting bars at the right side of primer sequences indicate primer combinations
4 based on the scaffolding analysis by TOPAAS. Redundant primer sequences are indicated in
5 red. A mouse-over function on the number of redundant blast hits will display alternative
6 priming positions and physical properties.
7

Blast Result of 046G10D00005_SP6

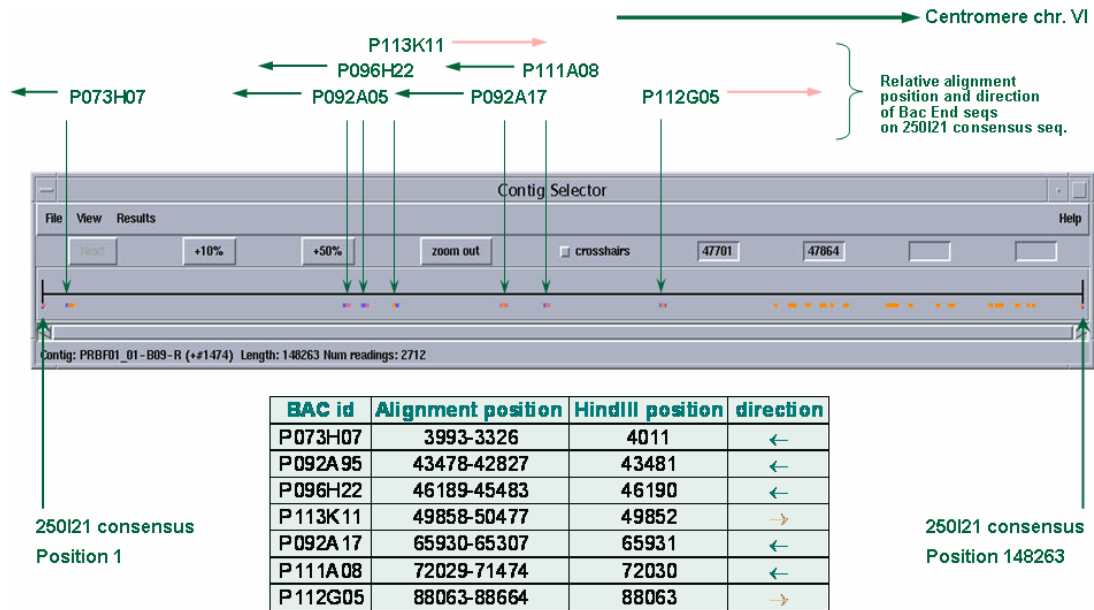
Program BLASTN Database SGN HindIII Tomato BAC Ends v.1.5 Database letters 76077097
Version 2.2.5 [Nov-16-2002] Database release (vector screened and Apr 14, 2005 4:47 PM Database seqs 126836
Query length 22710



8
9 **Fig. S3:** Schematic representation of a BlastN analysis of tomato contig P046G10D00005-
10 SP6 against the SGN BAC end sequence database. The linear sequence of the contig is
11 represented by horizontal green bars running from position 1 at the left site to position 22710
12 at the right site. Each BAC end hit is marked with a tick and positioned according to
13 homologous contig coordinates. The 15 most significant hits are displayed. Ticks are color
14 coded to indicate the level of significance (lower bar). At the left side, the BAC ID are

1 indicated. BAC end hit P046G10-SP6 is positioned at the right border of the BAC insert.
 2 P103N18-SP6 is positioned approximately 2 kb from the BAC insert end.

3

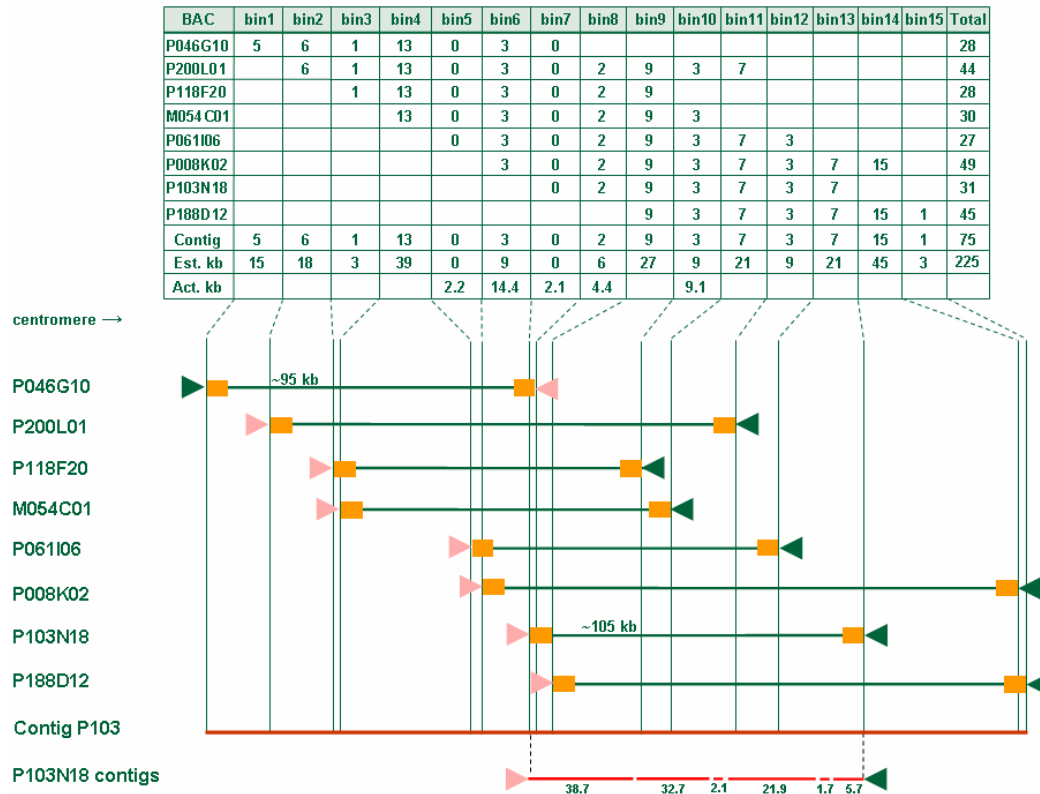


4



5

6 **Fig. S4:** Assembly of BAC ends onto the P250I21 consensus. In the contig selector panel the
 7 consensus of P250I21 is represented by the horizontal bar. Each assembled BAC position is
 8 indicated with a blue tick. Direction of the alignment is indicated by arrows above the panel.
 9 Coordinates for the alignment and upstream *Hind*III cloning position are shown in the table
 10 below the contig selector panel. In the join editor panel the unambiguous sequence alignment
 11 between seed BAC P250I21 and P073H07, and trace details are shown.



1
2
3
4
5
6
7
8
9
10

Fig. S5: BAC binning and physical map of contig P103. Each bin is defined as a domain in which a set of AFLP markers is shared between BACs. The amount of co-migrating markers indicated in the top table is used to estimate the order and size of the overlap. For each shared marker observed in gel an overlap portion of approximately 3kb is assumed. The assembly position of BAC end sequences (■) next to the T7 (▶) and SP6 (◀) region to the consensus of P046G10 have been used to calculate actual overlap sizes. BAC end positions are projected onto P103N18 contigs (—) for ordering.

Project name			
The file with the consensus data			<input type="button" value="Browse..."/>
The file with the readpair information			<input type="button" value="Browse..."/>
The file with the blast output			<input type="button" value="Browse..."/> <input type="checkbox"/> Perform BLAST on consensus in background Click to change variables
<input type="checkbox"/> Against non-redundant databases			
Program name	<input type="text" value="blastx"/>		
Public Database	<input type="text" value="nr"/>		
E-mail address	<input type="text"/>		
Send blast result by mail	<input type="checkbox"/>		
Other options for BLAST	<input type="text" value="-a 4"/>		
<input type="checkbox"/> Against BAC-ends			
Program name	<input type="text" value="megablast"/>		
Database	<input type="text" value="blastdb_bacends_trimmed_350"/>		
E-mail address	<input type="text"/>		
Send blast result by mail	<input type="checkbox"/>		
Other options for BLAST	<input type="text" value="-n T -a 4"/>		
The file with the nummer data			<input type="button" value="Browse..."/> <input type="checkbox"/> Perform genome alignment on consensus in background Click to change variables
<input type="checkbox"/> Using Mummer			
Reference sequence file	<input type="text" value="est_Solanum_total"/>		
Distance an alignment extension will attempt to extend poor scoring regions before giving up	<input type="text" value="200"/>		
Minimum length of a cluster of matches	<input type="text" value="65"/>		
Clustering fraction	<input type="text" value="0.12"/>		
Direction of alignment	<input type="text" value="forward and reverse"/>		
Minimum match of length	<input type="text" value="20"/>		
Maximum gap size between two adjacent matches	<input type="text" value="90"/>		
E-mail address	<input type="text"/>		
Send mummer result by mail	<input type="checkbox"/>		
<input type="checkbox"/> Using BLAT			
Reference sequence file	<input type="text" value="est_Solanum_total"/>		
Size of match that triggers an alignment	<input type="text" value="11"/>		
Spacing between files	<input type="text" value="11"/>		
One Off	<input type="checkbox"/>		
Number of file matches	<input type="text" value="2"/>		
Minimum score	<input type="text" value="30"/>		
Minimum sequence identity (in percent)	<input type="text" value="90"/>		
Size of maximum gap between files in a clump	<input type="text" value="2"/>		
Number of repetitions of a file allowed before it is marked as overused	<input type="text" value="1024"/>		
Maximum intron size	<input type="text" value="750000"/>		
Allow extension of alignment through large blocks of N's	<input type="checkbox"/>		
E-mail address	<input type="text"/>		
Send blat result by mail	<input type="checkbox"/>		
Leave entries blank if not used		<input type="button" value="Submit Query"/>	

1

2

3 **Fig. S6:** Web front-end for data upload and homology based alignments. Homology based
 4 alignments are either uploaded manually or provided by TOPAAS via two automated Blasts
 5 against the non redundant sequence database from NCBI and against the BAC end sequence
 6 database from SGN. The system provides for alignment of contigs against *Solanaceae* ESTs
 7 either with MUMmer or BLAT.

8

9

10

BAC id	size (kb)	contigs	TOPAAS links												gaps/ links	
			R		B		E		RB		RE		BE			RBE
RH123P09	131	4	1 (1)	0		0		0		1 (1)	gb BG132480 gb BM110288	1 (1)	dbj BAD08898 gi AI775618	0		3/3
SH196	72	11	10 (9)	0		0		0		0		0		0		9/10
RH011D17	132	6	2 (2)	2 (2)	gb AF220602 gb S51460	1 (1)	gb BM407793	0		0		0		0		5/5
P073H07	130	18	7 (4)	0		0		1 (1)	gb AAL68851	1 (1)	gb BI927628 gb BG597118	0		1 (1)	pir B84710 gb B1176827 gb BQ509205	7/10
P103N18	105	6	4 (4)	0		0		1 (1)	gb AAC72337	0		0		0		5/5
P250I21	148	1	-	-		-		-		-		-		-		n.d.
P046G10	90	8 (7)	-	-		-		-		-		-		-		n.d.

1

2 **Table T1:** Link analysis by TOPAAS for potato and tomato BACs. Tomato BAC IDs are
3 indicated with a prefix P, and potato BACs have a prefix RH or SH. For each BAC the insert
4 size and the amount of contigs remaining after shotgun assembly is given. Linkage result is
5 represented by the number of contig pairs linked with gap flanking read pairs (R), gap
6 bridging BlastX hits (B), EST gap spanning alignments (E) and combinations thereof. The
7 gap closure result for each link type per BAC is indicated between brackets. References for
8 Blast and EST matches are indicated with Genbank, or PIR database accession numbers. The
9 closing efficiency is shown as the number of closed gaps over the number of predicted contig
10 links per BAC. Link analysis was not determined for P250I21 and P046G10.

11