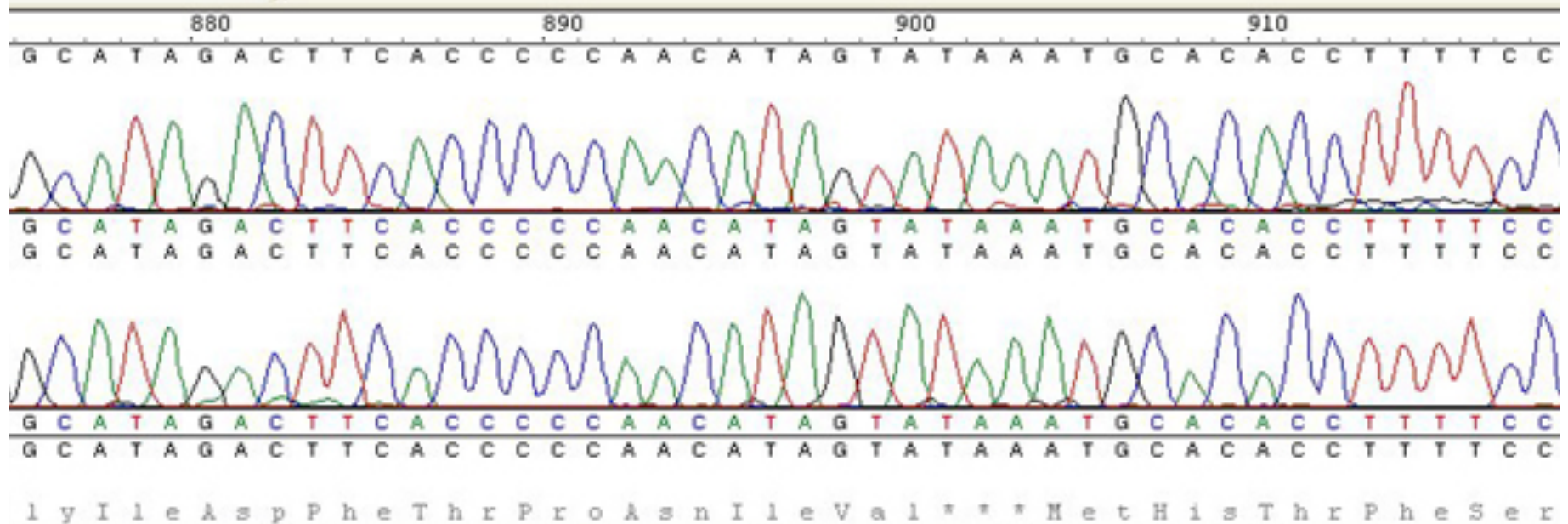




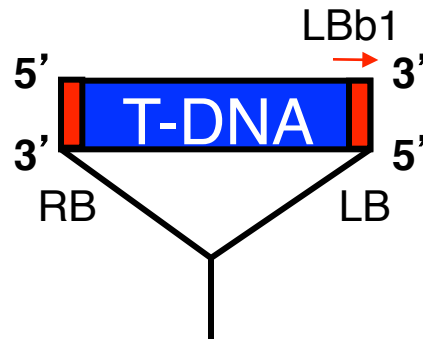
Bioinformatics Session

I. Determine T-DNA Insertion Site

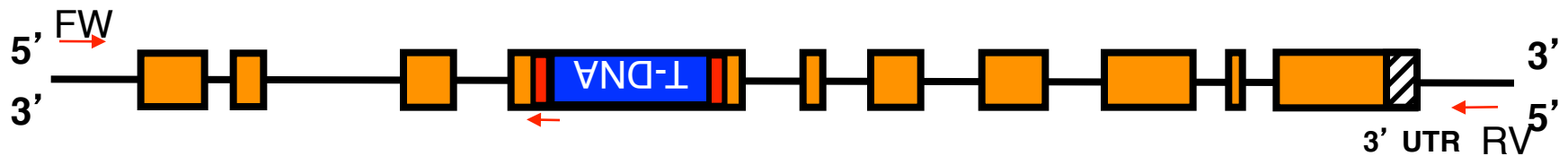
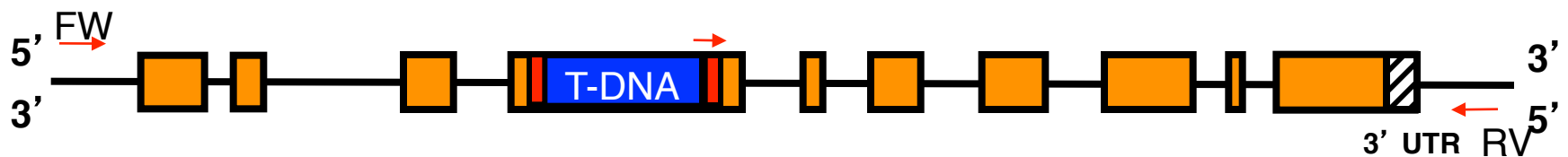
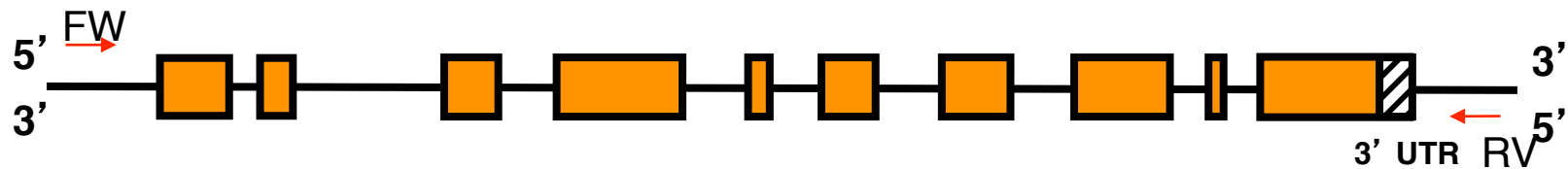
HC70AL
Spring 2011

April 14, 2011

Where is the T-DNA Inserted Within the Gene?



?????



Determining T-DNA Insertion Site - PART I

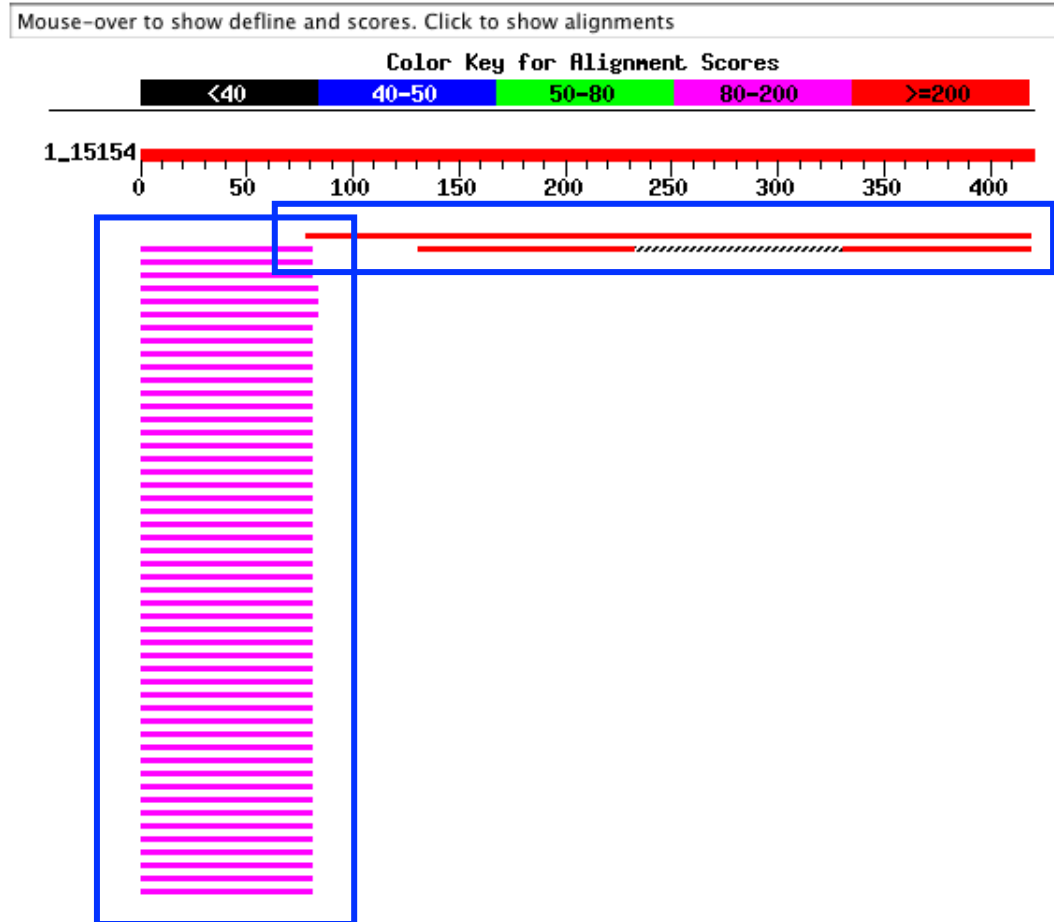
Sequence results

1

420

Unknown Sequence

Distribution of 105 Blast Hits on the Query Sequence



Determining T-DNA Insertion Site - PART II

Sequences producing significant alignments:			Score (bits)	E Value	
gi 20197427 gb AC005700.3 	Arabidopsis thaliana chromosome ...	615	e-173		
gi 30685281 ref NM_128796.2 	Arabidopsis thaliana homeobox-...	202	4e-49		
gi 21999149 gb AY096770.1 	Populus tremula x Populus tremul...	163	3e-37		
gi 21999147 gb AY096768.1 	Populus tremula x Populus tremul...	163	3e-37		
gi 21999152 gb AY096773.1 	Populus tremula x Populus tremul...	163	3e-37		
gi 22417148 gb AF531263.1 	Populus tremula transgenic isola...	153	3e-34		
gi 21999138 gb AY096759.1 	Populus tremula transgenic isola...	153	3e-34		
gi 10639841 emb AJ296082.1 PTR296082	Populus tremula transg...	153	3e-34		
gi 38455225 gb AY456904.1 	Binary vector pRE1, complete seq...	149	5e-33		
gi 22773774 gb AY141042.1 	Binary vector pGV4945 kanamycin ...	149	5e-33		
gi 30027630 gb AY218787.1 	Cloning vector pAJ1, complete se...	149	5e-33		
gi 29569730 gb AY234331.1 	Binary vector pLH7500, complete ...	149	5e-33		
gi 29569724 gb AY234330.1 	Binary vector pLH7000, complete ...	149	5e-33		
gi 29569718 gb AY234329.1 	Binary vector pLH6500, complete ...	149	5e-33		
gi 29569712 gb AY234328.1 	Binary vector pLH6000, complete ...	149	5e-33		
gi 29569706 gb AY234327.1 	Binary vector pLH5000, complete ...	149	5e-33		
gi 29569701 gb AY234326.1 	Binary vector pLHBA, complete se...	149	5e-33		
gi 29569696 gb AY234325.1 	Binary vector pLHAB, complete se...	149	5e-33		

Determining T-DNA Insertion Site - PART III

```
>gi|20197427|gb|AC005700.3|   Arabidopsis thaliana chromosome 2 clone T32F6 map TEn5, complete  
sequence  
Length = 100424
```

```
Score = 615 bits (310), Expect = e-173  
Identities = 332/342 (97%)  
Strand = Plus / Minus
```

```
Query: 79      tatactataacatatattaggtcaggtgtgtaattaatgacctgtatttgcaaaatctgt 138  
            |||  
Sbjct: 44094  tatactataacatatattaggtcaggtgtgtaattaatgacctgtatttgcagaatctgt 44035
```

```
Query: 139     ggtcaacaatgtttgccggaattgttggtagagccaggactcatgaacagataatggctg 198  
            |||  
Sbjct: 44034  ggtcaacaatgtttgccggaattgttggtagagccaggactcatgaacagataatggctg 43975
```

```
Query: 199     atgctgctggaaacttcaatggaaatctccaaatagtaagtcttctacactaatatagag 258  
            |||  
Sbjct: 43974  atgctgctggaaacttcaatggaaatctccaaatagtaagtcttctacactaatatagag 43915
```

```
Query: 259     ttatacaaagnnnnnnnncaatacttgtttgatataataaaagtttaaccaattaatattg 318  
            |||  
Sbjct: 43914  ttatacaaagaaaaaaaacaatacttgtttgatataataaaagtttaaccaattaatattg 43855
```

```
Query: 319     ttcttattcatagatgagtgtgagtaccaagtgtttcccccgtagtcacaacccgcga 378  
            |||  
Sbjct: 43854  ttcttattcatagatgagtgtgagtaccaagtgtttcccccgtagtcacaacccgcga 43795
```

```
Query: 379     aagctacttcgtccgctactgaaagcaacaaggagaggggtt 420  
            |||  
Sbjct: 43794  aagctacttcgtccgctactgtaagcaacaaggagaggggtt 43753
```

Examination of Chromosome Clone With Identical Sequence

```
gene      complement(42398..45830)
          /gene="At2g32370"
          /note="synonym: T32F6.11"
mRNA      complement(join(<42398..42751,42836..42925,
          43013..43275,43370..43541,43641..43841,43940..44041,
          44144..44566,44707..44895,45344..45461,45577..>45830))
          /gene="At2g32370"
CDS       complement(join(42398..42751,42836..42925,43013..43275,
          43370..43541,43641..43841,43940..44041,44144..44566,
          44707..44895,45344..45461,45577..45830))
          /gene="At2g32370"
          /codon_start=1
          /product="putative homeodomain transcription factor"
          /protein_id="AAC69941.1"
          /db_xref="GI:3831459"
          /translation="MSQSNMVPVANNGDNNNDNENNNNNNNNNGGTDNTNAGNDSGDQD
          FDSGNTSSGNHGEGLGNNQAPRHKKKKYNRHTQLQISEMEAFFRECPHPDDKQRYDLS
          AQLGLDPVQIKFWFQNKRTQNKNOQERFENSELRNLNNHLRSENQRLREAIHQALCPK
          CGGQTAIGEMTFEEHHLRILNARLTTEEIKQLSVTAEKISRLTGIPVRSHPRVSPPNPP
          PNFEFGMGSKGNVGNHSRETTGPADANTKPIIMELAFGAMEELLVMAQVAEPLWMGGF
          NGTSLALNLDEYEKTFRTGLGPRLGGFRTEASRETALVAMCPTGIVEMLMQENLWSTM
          FAGIVGRARTHEQIMADAAGNFNGNLQIMSAEYQVLSPLVTTRESYFVRYCKQQGEGL
          WAVVDISIDHLLPNINLKCRRRPSGCLIQEMHSGYSKVTWVEHVEVDDAGSYSIFEKL
          ICTGQAFANRWVGT LVRQCERISSILSTDFQSVD SALTNHGKMSMLKIAERIARTFF
          AGMTNATGSTIFSGVEGEDIRVMTMKS VNDPGKPPGVIICAATSFWL PAPNTVFD FL
          REATHRHNDVLCNGEMMHKIAEITNGIDKRNCASLLRHGHTSKSKMMIVQETSTDPT
          ASFVLYAPVDMTSM DITLHGGGDPDFVVILPSGFAIFPDGTGKPGGKEGGSLLTISFQ
          MLVESGPEARLSVSSVATTENLIRTTVRRIKDLFPCQTA"
```

Determining T-DNA Insertion Site - PART IV

```
>gi|30685281|ref|NM\_128796.2| U Arabidopsis thaliana homeobox-leucine zipper family protein /  
lipid-binding START domain-containing protein (At2g32370)  
mRNA, complete cds  
Length = 2612
```

```
Score = 202 bits (102), Expect = 4e-49  
Identities = 102/102 (100%)  
Strand = Plus / Plus
```

```
Query: 132 aatctgtggtcaacaatgtttgccggaattgttggtagagccaggactcatgaacagata 191  
          ||||||||||||||||||||||||||||||||||||||||||||||||||||  
Sbjct: 985 aatctgtggtcaacaatgtttgccggaattgttggtagagccaggactcatgaacagata 1044
```

```
Query: 192 atggctgatgctgctggaaacttcaatggaaatctccaaata 233  
          ||||||||||||||||||||||||||||||||||||||||||||  
Sbjct: 1045 atggctgatgctgctggaaacttcaatggaaatctccaaata 1086
```

```
Score = 168 bits (85), Expect = 6e-39  
Identities = 88/89 (98%)  
Strand = Plus / Plus
```

```
Query: 332 atgagtgctgagtaccaagtgctttccccgctagtcacaacccgcgaaagctacttcgtc 391  
          ||||||||||||||||||||||||||||||||||||||||||||  
Sbjct: 1087 atgagtgctgagtaccaagtgctttccccgctagtcacaacccgcgaaagctacttcgtc 1146
```

```
Query: 392 cgctactgaaagcaacaaggagagggttt 420  
          ||||||| ||||||||||||||||  
Sbjct: 1147 cgctactgtaagcaacaaggagagggttt 1175
```


Determining T-DNA Insertion Site - PART V

```
 >gi|29569706|gb|AY234327.1| Binary vector pLH5000, complete sequence  
Length = 8958
```

```
Score = 149 bits (75), Expect = 5e-33  
Identities = 82/83 (98%), Gaps = 1/83 (1%)  
Strand = Plus / Minus
```

```
Query: 1      ttctttttctcc-tattgaccatcatactcattgctgatccatgtagatttcccgacat 59  
            |||  
Sbjct: 8746   ttctttttctccatattgaccatcatactcattgctgatccatgtagatttcccgacat 8687
```

```
Query: 60     gaagccatttacaattgaatata 82  
            |||  
Sbjct: 8686   gaagccatttacaattgaatata 8664
```

```
 >gi|29569701|gb|AY234326.1| Binary vector pLHBA, complete sequence  
Length = 7674
```

```
Score = 149 bits (75), Expect = 5e-33  
Identities = 82/83 (98%), Gaps = 1/83 (1%)  
Strand = Plus / Minus
```

```
Query: 1      ttctttttctcc-tattgaccatcatactcattgctgatccatgtagatttcccgacat 59  
            |||  
Sbjct: 7462   ttctttttctccatattgaccatcatactcattgctgatccatgtagatttcccgacat 7403
```

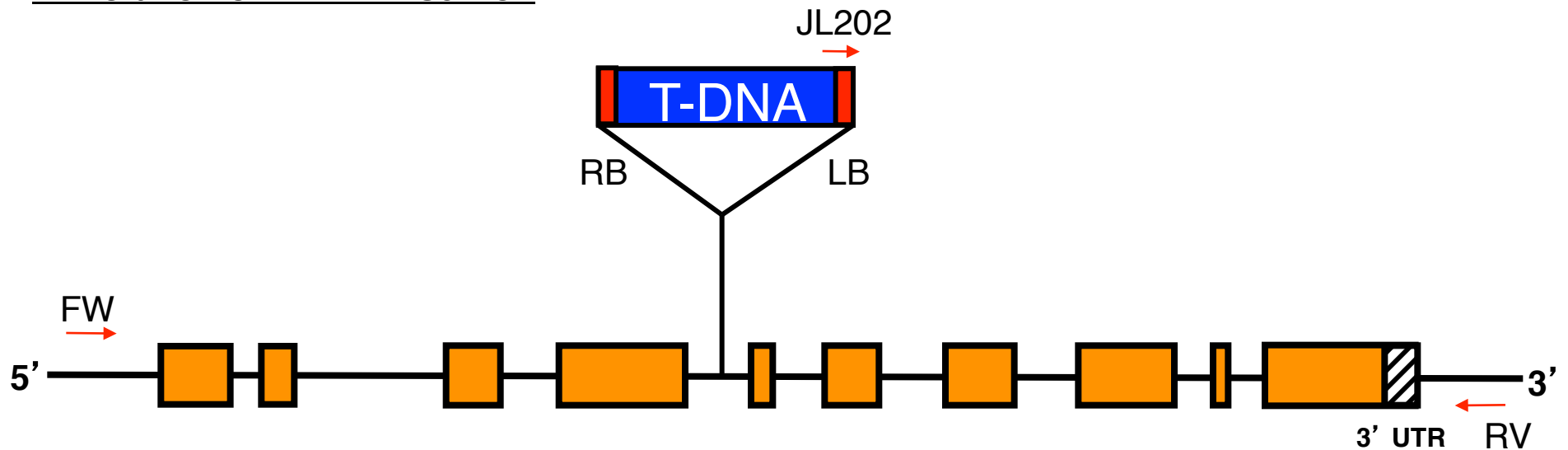
```
Query: 60     gaagccatttacaattgaatata 82  
            |||  
Sbjct: 7402   gaagccatttacaattgaatata 7380
```


T-DNA Inserted in an Intron Within the Gene!!

Sequence results



Annotation of T-DNA Insertion



Identifying T-DNA Insertion Site Using SeqViewer

Chromosome 2 start: 13749001 length: 10000

http://arabidopsis.org/servlets/sv?action=seq&pid=0&option=0&option2=0&on=0

Lab Bioinformatics Reagents FileMaker Db LINUX Essentials Macinstruct MIT Courses BioInfoBank Library Cancer Resea...ray Service Reactome TV LINKS

Chromosome 2 start: 1374...

tair ? SeqViewer Nucleotide View SeqViewer Home | Release Note

ATC = Translational Start/Stop
ATGC = Exon
atgc = Annotation on other strand Hide

atgc = UTR
atgc = Intron

Choose strand to view:
Forward Strand

Choose objects to be highlighted:
Genes

Clicking on the arrows will shift the view by 5 kb

10 20 30 40 50 60 70 80 90 100 Gene(s) + Marker(s)

13749001 caaaatatttttggggttaatttaaatagttctatatatatataatttcattttgtttgggttgattattttgttaattctctactggtgggtttttttatttcc
aagtttccaactatcatcctcaattactttgtctttttttttttgtctgtttctcttttatctataagatcttctagaaaaagaaaaataagataaaaa
13749201 attggttgacttggaagcaaaccttttgggtgcaggaatttagtgaaagagtgtagtgtttgtctgactgcaaaaaatccaaatcttggaaaaaa
tcctttttgattcttggtaactcaagaagaaaaagaaaaagaaaccttttcacaaaccctaatttcttagatcttatgaattcttatcattttatatc
13749401 cctctaatttgatgtaactcatgatattttctatttcaactctcagatcttttagaatcccaaagaaaagaaaacaacaacaaaaATCTCTCAGTCAAACA
TGGTACCAGTGGCTAACACGGAGACAACAACGACAACGAAAAACAACAACAACAACAACAATGGTGGAACGTGACAACTAATGCTGGAAATGA
13749601 TTCTGGAGATCAAGATTTCGACAGTGGGAATACCTCAAGTGGCAATCATGGAGAAGGGTTGGGAAACAATCAAGCTCCTCGTCATAAGAAGAAAAATAC
AATCGTCACACCCAACCTCAGATTTCGGAGATGGAAGCgtaatttttcttcattttcttaataactcaacttctatctagctagatcagtaactaagtta
13749801 tttataatcatttcacatctctccttttttctttttctttttatgtgttggaatagTTCTCAGAGAGTGTCTCACCAGATGACAAACAAGGTACGACC
TTAGCGCTCAATGGGATTGGACCTGTTTCAGATCAAATTCTGGTTCAGAACAAACGCACTCAAACAAGgtacgtatatatgtaataaattttctcaca
13750001 atctcacacatatattttctgtatgtatatataggcatgagttggcgttgaataattcttggatatataagcttagagtagtggttccacatatatattatt
tgttgtaattgtagtaagatttttgtgaaataaatgaaatcaagtgtttttttttgtgtcttttttccgtaagagaatattataataacattttgaaac
13750201 gtgtggaacaacacacagtttttgcacaggtgtatatatctgtagacgaagtgtgtgtgtttgtgacgcaattaatacgcacatttggtttattctttatt
cacttgaaacacacacagctcatttttaatttgcgtgcttccatatccaatgataacgacttaattagttgttttttcagtttaagttgaaatataactaagaa
13750401 cttttgttttgttttgcagAATCAACAAGAACGCTTTGAGAACTCAGAACTTCGGAATCTGAACAACACCTTAGGTCTGAAAATCAGCGGTTACGAGAA
GCTATTCATCAAGCCTTATGCCCTAAGTGTGGAGGCCAACTGCAATTGGCGAAATGACCTTCGAAGAGCACCATCTTCGCATCCTCAACGCTCGTTTGA
13750601 CTGAAGAGgtactagattgaactagagcattttctattatgtaaatgtaaatgttcataacaattgcacattgactatttcaatgaatatgaaagtaaca
cttcaaaaatgatttaacaagttgtattttctgttttgggttatagATCAAGCAACTTCCGTGACAGCGGAAAAGATATCAAGGCTTACGGGGATAC
13750801 CAGTAAGGAGCCATCCCGTGTGTCTCCTCCTAATCCTCCTCCAAATTCGAGTTCGGGATGGGATCTAAGGGAATGTCTGGAAACCACTCGAGGGAAAC
CACTGGACCTGCAGATGCTAATACCAAGCCGATCATCATGGAGTTGGCATTTCGAGCCATGGAGGAGCTCTTGGTGATGGCTCAAGTGGCTGAACCACTG
13751001 TGGATGGGAGGATTTAATGGCACTAGCTTAGCTTTGAACTTGGATGAATACGAAAAGACGTTTCGCACGGGTCTCGGTCTAGACTTGGCGGGTTTCGAA
CCGAGGCATCCAGGGAACCTGCACTCGTGGCAATGTGTCTACTGGCATTGTTGAAATGCTCATGCAAGAGgtaaattaatcaatcaattatatgatca
13751201 taatttaactccttttatatagataactatacatatattaggtcaggtgtgttaattaatgacctgtattttgcagAATCTGTGGTCAACAATGTTTGGCGGA
ATTGTTGGTAGAGCCAGGACTCATGAACAGATAATGGCTGATGCTGCTGGAACTTCAATGGAAATCTCCAATAgtaagtcttctacactaatatagag

Identifying T-DNA Insertion Site Using SeqViewer

[illegible]

SALK T-DNA Lines

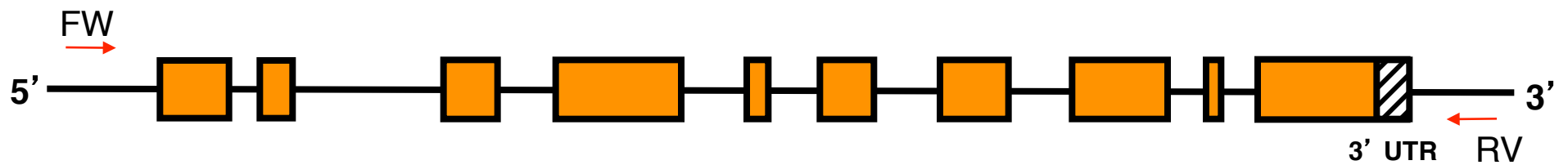
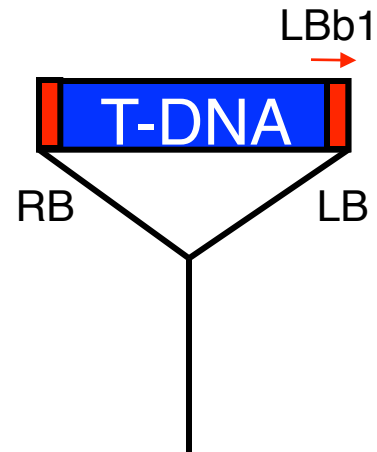
A Collection of T-DNA Mutant Lines



<http://signal.salk.edu/cgi-bin/tdnaexpress>

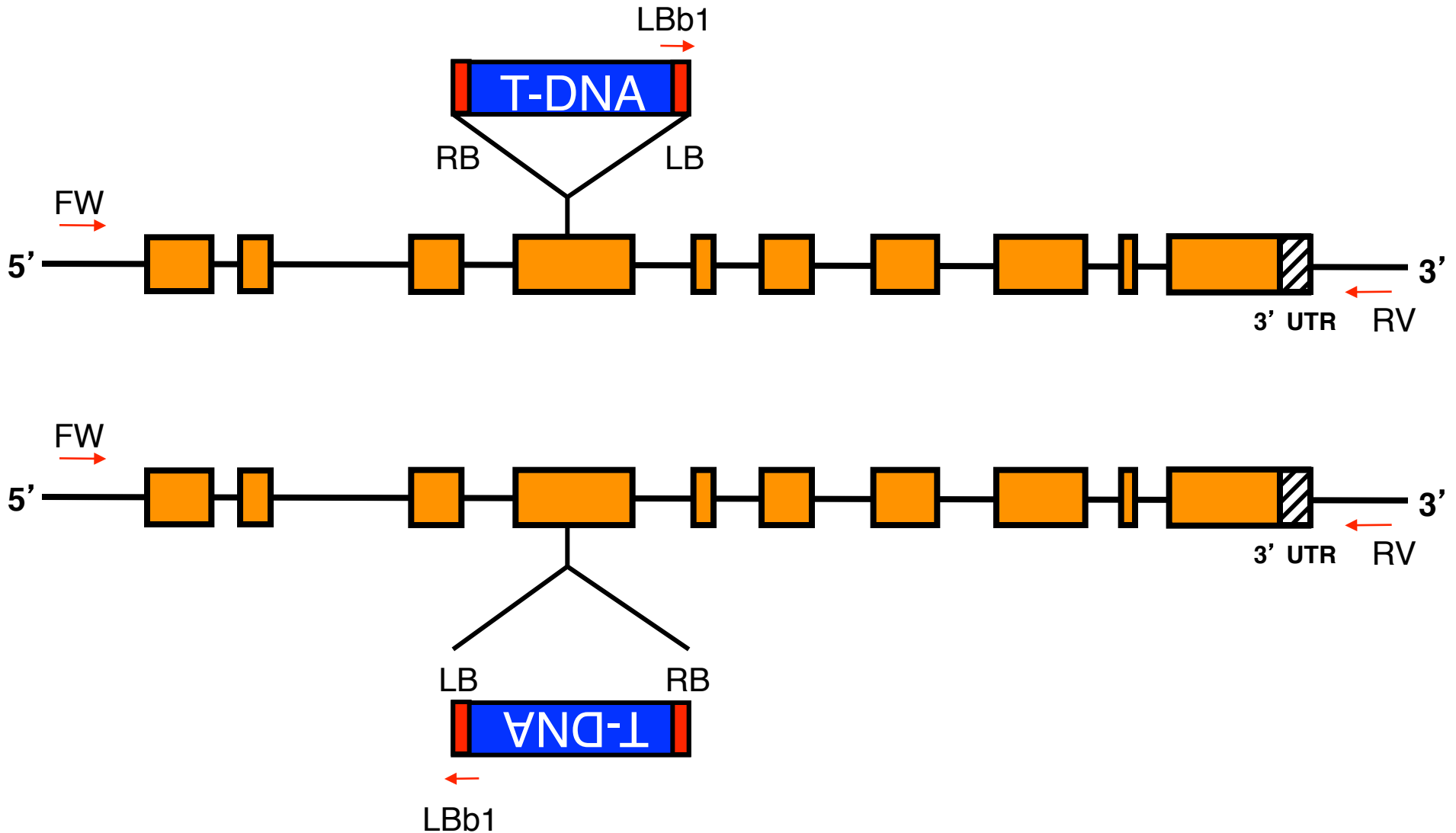
SALK T-DNA Lines

How Were T-DNA Lines Created?



SALK T-DNA Lines

How Were T-DNA Lines Created?



SALK T-DNA Lines

A Collection of T-DNA Mutant Lines

T-DNAexpress : The SIGnAL Arabidopsis Gene Mapping Tool

http://signal.salk.edu/cgi-bin/tdnaexpress

Lab HC70A Bioinformatics Arabidopsis Phytome Scholar Softpedia

SIGnAL Funded by the National Science Foundation
Salk Institute Genomic Analysis Laboratory

SIGnAL "T-DNA Express" Arabidopsis Gene Mapping Tool (Apr. 8, 2005)

Arabidopsis thaliana [AGI V5]
chr1 1 - 25001

out

Please note: We are using AGI V5 pseudo-molecules (Feb-19-2004)/annotation (Feb-23-2005, revised Mar-17-2005).

1. Edit Preference:
[Click to edit your preference for T-DNA Express.](#)

4. Blast:
Program: E-value: Hits:

SALK T-DNA Lines

A Collection of T-DNA Mutant Lines

T-DNAexpress : The SIGnAL Arabidopsis Gene Mapping Tool

http://signal.salk.edu/cgi-bin/tdnaexpress

Lab ▾ HC70A ▾ Bioinformatics ▾ Arabidopsis ▾ Phytome ▾ Scholar ▾ Softpedia

[PY11](#) [GSLT cDNA](#) [RIKEN EST](#) [PY11a01010](#) [RATM13-3326-1_G](#) [SALK_092277](#)

Please note: We are using AGI V5 pseudo-molecules (Feb-19-2004)/annotation (Feb-23-2005, revised Mar-17-2005).

1. Edit Preference:

[Click to edit your preference for T-DNA Express.](#)

2. Search:

Gene name: for example: [At1g01010](#)

or Function: for example: [auxin](#) (Case sensitive)

or cDNA/T-DNA: for example: [SALK_003854](#), [R10047](#)

4. Blast:

Program: E-value: Hits:

Cut and paste your sequence into here.

File : no file selected

3. Locate:

Position: chr1

5. Multiple Search

Search by Gene/Affy/GO/Sequence and plain text return only.

6. SIGnAL iSect Tools

to design primers for T-DNA or your sequences, to retrieve sequences and to compare two lists.

ANNOUNCEMENT : **

- We requested that users include this reference (below) in publications describing the use of SSP gold standard cDNA ORF clones obtained from ABRC. Thanks!
 Kayoko Yamada, Jun Lim, Joseph M. Dale, Huaming Chen, Paul Shinn, Curtis J. Palm, Audrey M. Southwick, Hank C. Wu, Christopher Kim, Michelle Nguyen, Paul Pham, Rosa Cheuk, George Karlin-Newmann, Shirley X. Liu, Bao Lam, Hitomi Sakano, Troy Wu, Guixia Yu, Molly Miranda, Hong L. Quach, Matthew Tripp, Charlie H. Chang, Jeong M. Lee, Mitsue Toriumi, Marie M. H. Chan, Carolyn C. Tang, Courtney S. Onodera, Justine M. Deng, Kenji Akiyama, Yasser Ansari, Takahiro Arakawa, Jenny Banh, Fumika Banno, Leah Bowser, Shelsea Brooks, Piero Carninci, Qimin Chao, Nathan Choy, Akiko Enju, Andrew D. Goldsmith, Mani Gurjal, Nancy F. Hansen, Yoshihide Hayashizaki, Chanda Johnson-Hopson, Vickie W. Hsuan, Kei Iida, Meagan Karnes, Shehnaz Khan, Eric Koesema, Junko Ishida, Paul X. Jiang, Ted Jones, Jun Kawai, Asako Kamiya, Cristina Meyers, Maiko Nakajima, Mari Narusaka, Motoaki Seki, Tetsuya Sakurai, Masakazu Satou, Racquel Tamse, Maria Vaysberg, Erika K. Wallender, Cecilia Wong, Yuki Yamamura, Shialou Yuan, Kazuo Shinozaki, Ronald W. Davis, Athanasios Theologis, and Joseph R. Ecker (2003) **Empirical Analysis of Transcriptional Activity in the Arabidopsis Genome Science 302: 842-846.**
[\[Abstract \]](#) [\[Full Text \]](#) [\[Supporting data \]](#) [\[Supplementary GEO Analysis Files \]](#)
- We requested that users include this reference (below) in publications describing the use of Salk lines obtained from ABRC or NASC. Thanks!
 José M. Alonso, Anna N. Stepanova, Thomas J. Leisse, Christopher J. Kim, Huaming Chen, Paul Shinn, Denise K. Stevenson, Justin Zimmerman, Pascual Barajas, Rosa Cheuk, Carmelita Gadrinab, Collen Heller, Albert Jeske, Eric Koesema, Cristina C. Meyers, Holly Parker, Lance Prednis, Yasser Ansari, Nathan Choy, Hashim Deen, Michael Geralt, Nisha Hazari, Emily Hom, Meagan Karnes, Celene Mulholland, Ral Ndubaku, Ian Schmidt, Plinio Guzman, Laura Aguilar-Henonin, Markus Schmid, Detlef Weigel, David E. Carter, Trudy

A Collection of T-DNA Mutant Lines

