FIGURE 1

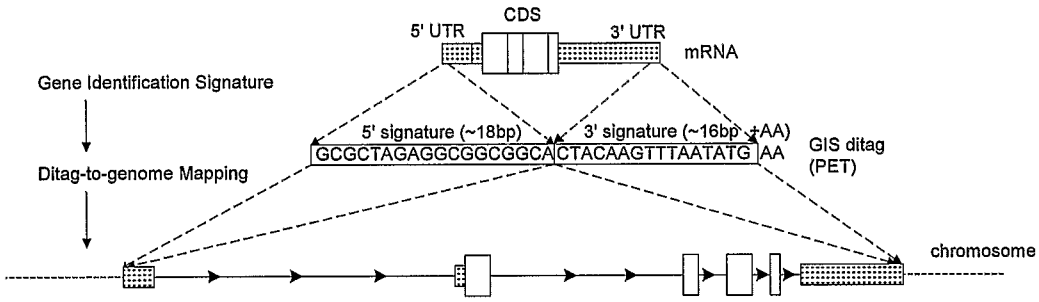

CDS
5' UTR
3' UTR
mRNA
Gene Identification Signature
5' signature (~18bp)
3' signature (~16bp +AA)
GCGCTAGAGGCGGCGGCA
CTACAAGTTTAATATG
AA
GIS ditag
(PET)
Ditag-to-genome Mapping
chromosome

FIGURE 2

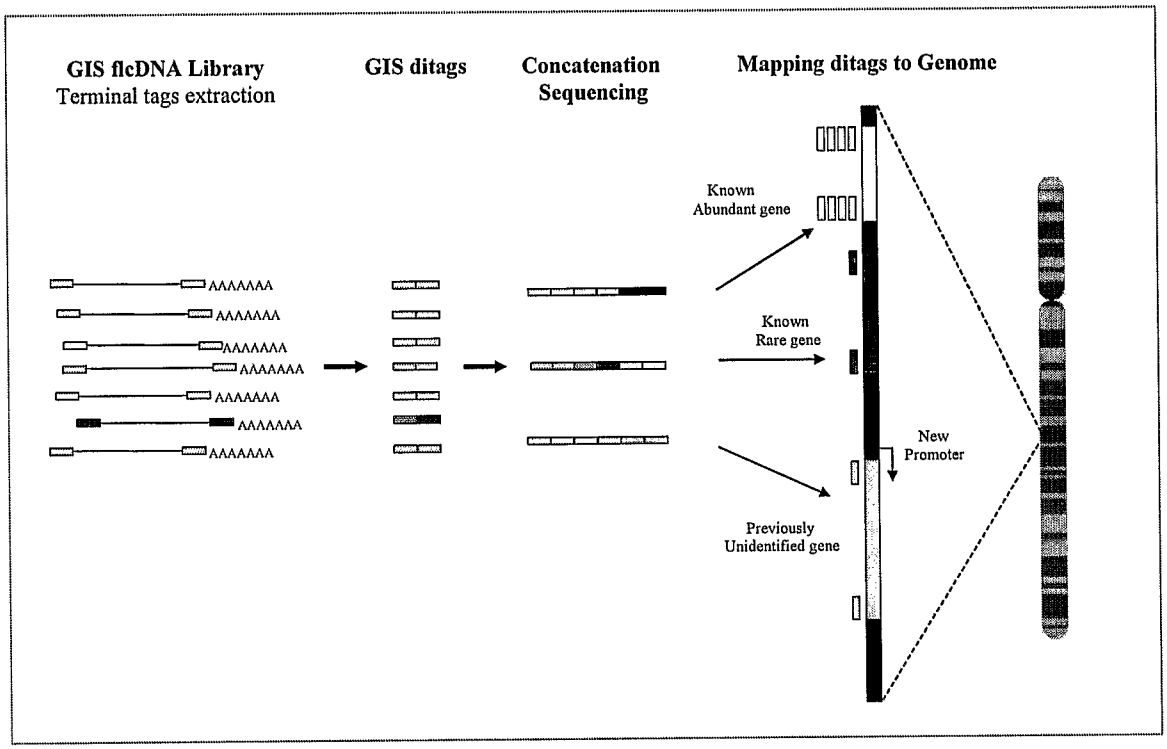

GIS flcDNA Library
Terminal tags extraction
GIS ditags
Concatenation
Sequencing
Mapping ditags to Genome
Known
Abundant gene
Known
Rare gene
New
Promoter
Previously
Unidentified gene

FIGURE 3

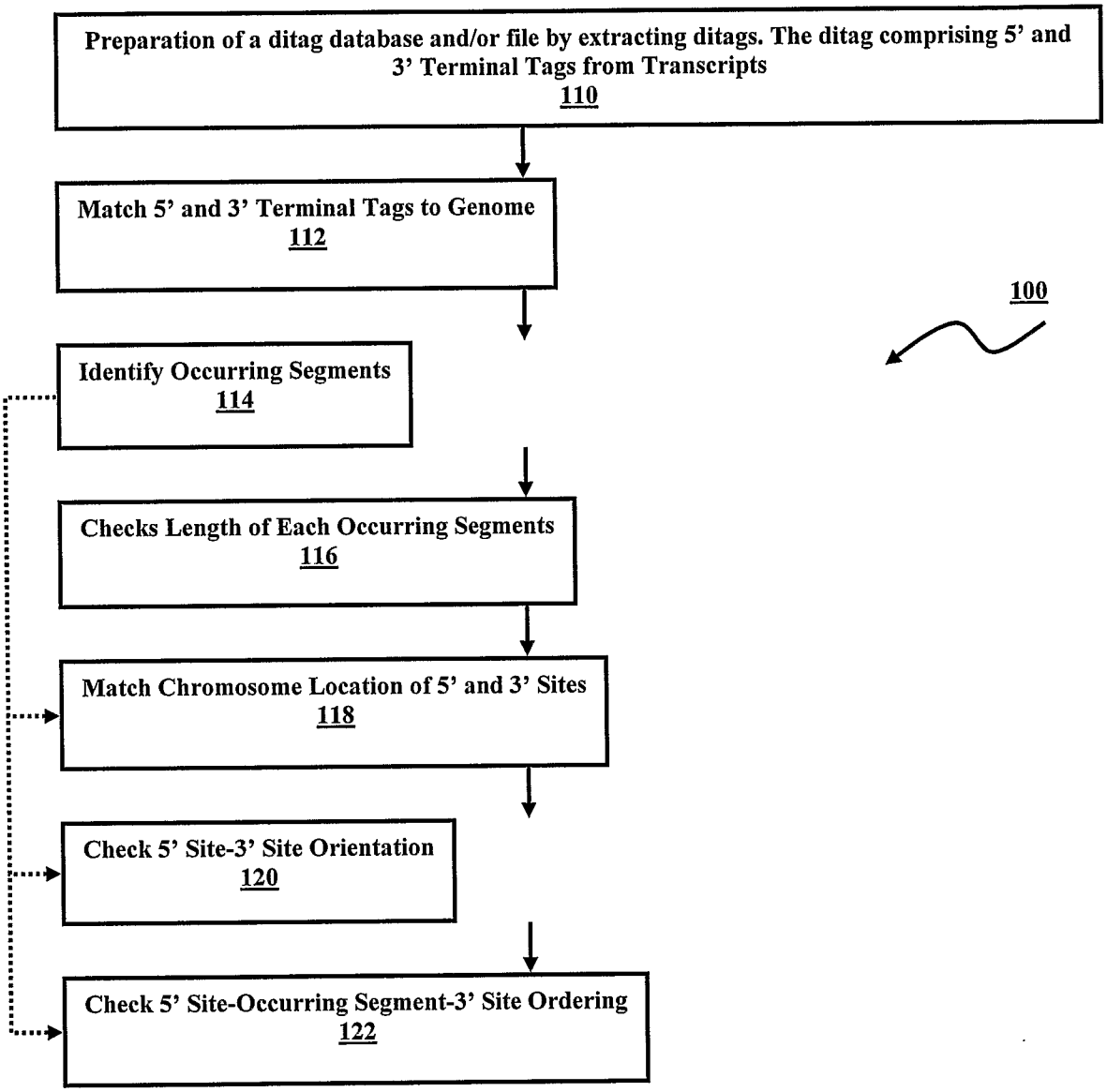

Preparation of a ditag database and/or file by extracting ditags. The ditag comprising 5' and 3' Terminal Tags from Transcripts
110
Match 5' and 3' Terminal Tags to Genome
112
100
Identify Occurring Segments
114
Checks Length of Each Occurring Segments
116
Match Chromosome Location of 5' and 3' Sites
118
Check 5' Site-3' Site Orientation
120
Check 5' Site-Occurring Segment-3' Site Ordering
122

FIGURE 4

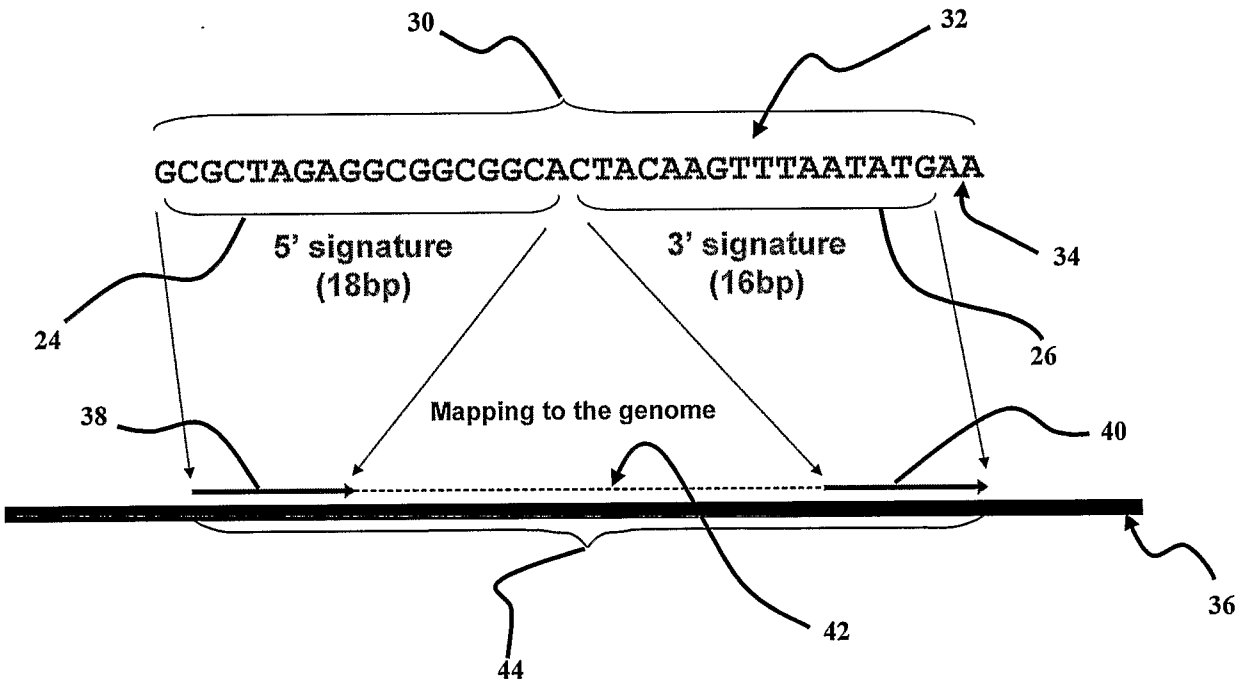

30
32
GCGCTAGAGGCGGCGGCACTACAAGTTTAATATGAA
5' signature
(18bp)
3' signature
(16bp)
34
24
26
38
Mapping to the genome
40
36
42
44

FIGURE 5

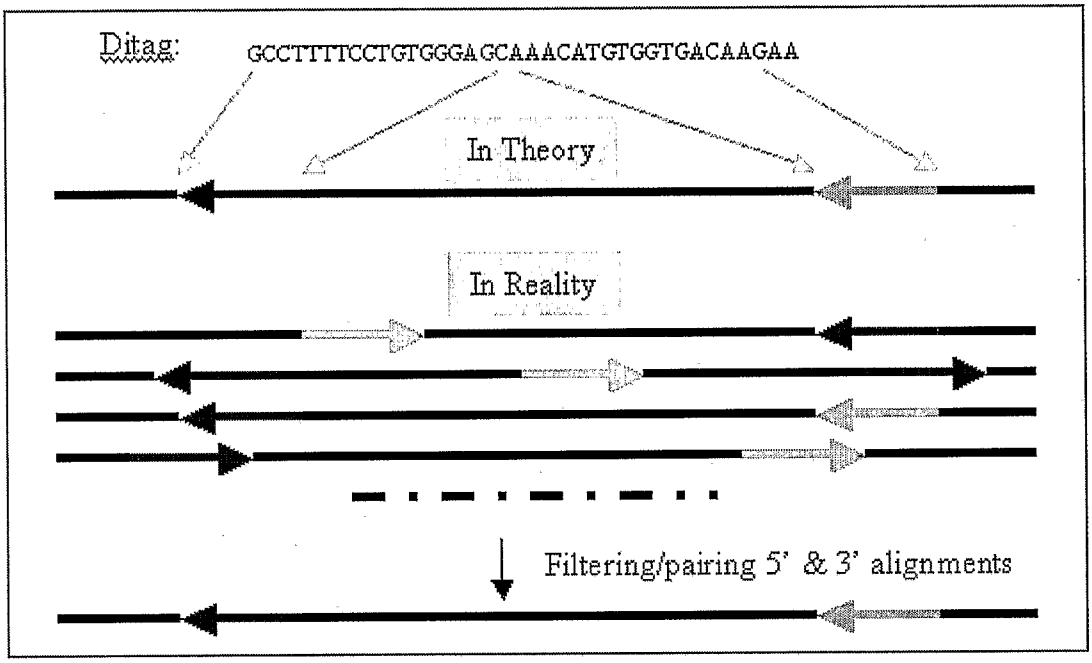

Ditag:
GCCTTTTCCTGTGGGAGCAAACATGTGGTGACAAGAA
In Theory
In Reality
Filtering/pairing 5' & 3' alignments

FIGURE 6

Algorithm Find_Sites (5' sites)
Input: A DNA segment P[1..m], a genome G[1..n], and a threshold $minlength_5$ of minimum contiguous exact match.
Output: A set of 5' sites.

for i = $minlength_5$ downto 1 **do**
Given the occurrence range of P[i+1..$minlength_5$], we find the occurrence range of P[i..$minlength_5$] by Backward search (Lemma 2).
Extend and report all the occurrences of P[1..$minlength_5$]
for i = 2 to $boundary_5$ - $minlength_5$ + 1 **do**
for j= $minlength_5$+1 to i+$minlength_5$-1 **do**
Given the occurrence range of P[i..j-1], we find the occurrence range of P[i..j] by Forward search (Lemma 1).
Extend and report all occurrences of P[i..i+$minlength_5$-1]

End Find_Sites (5' sites)

FIGURE 7

Algorithm Match_Sites_1
Input: A set of 5' sites and 3' sites.
Output: A set of occurring segments (feasible gene locations).

Sort the 5' sites in ascending order and let A[1], A[2], ..., A[p] be the list of 5' sites.
Sort the 3' sites in ascending order and let B[1], B[2], ..., B[q] be the list of 3' sites.
Let i=1 and j=1.
while (i ≤ p and j ≤ q) **do**
if (A[i]>B[j]) **then**
j = j + 1;
elseif (B[j]-A[i] ≤ *maxlength*)
report (A[i], B[j]) as an occurring segment and j=j+1;
else
i = i + 1;

End Match_Sites_1

FIGURE 8

Algorithm Match_Sites_2
Input: A set of 5' sites and 3' sites.
Output: A set of occurring segments.

Let i=1 and j=1.

if ($count_5$ << $count_3$) **then**
while (i ≤ p) **do**
    find 3' site located within *maxlength* downstream from A[i]
    report found 3' site and A[i] as an occurring segment
    i = i + 1;;
elseif ($count_5$ >> $count_3$)
while (j ≤ q) **do**
    find 5' site located within *maxlength* upstream from B[j]
    report found 5' site and B[j] as an occurring segment
    j = j + 1;;
else
    try Match_Sites_1

End Match_Sites_2

FIGURE 9

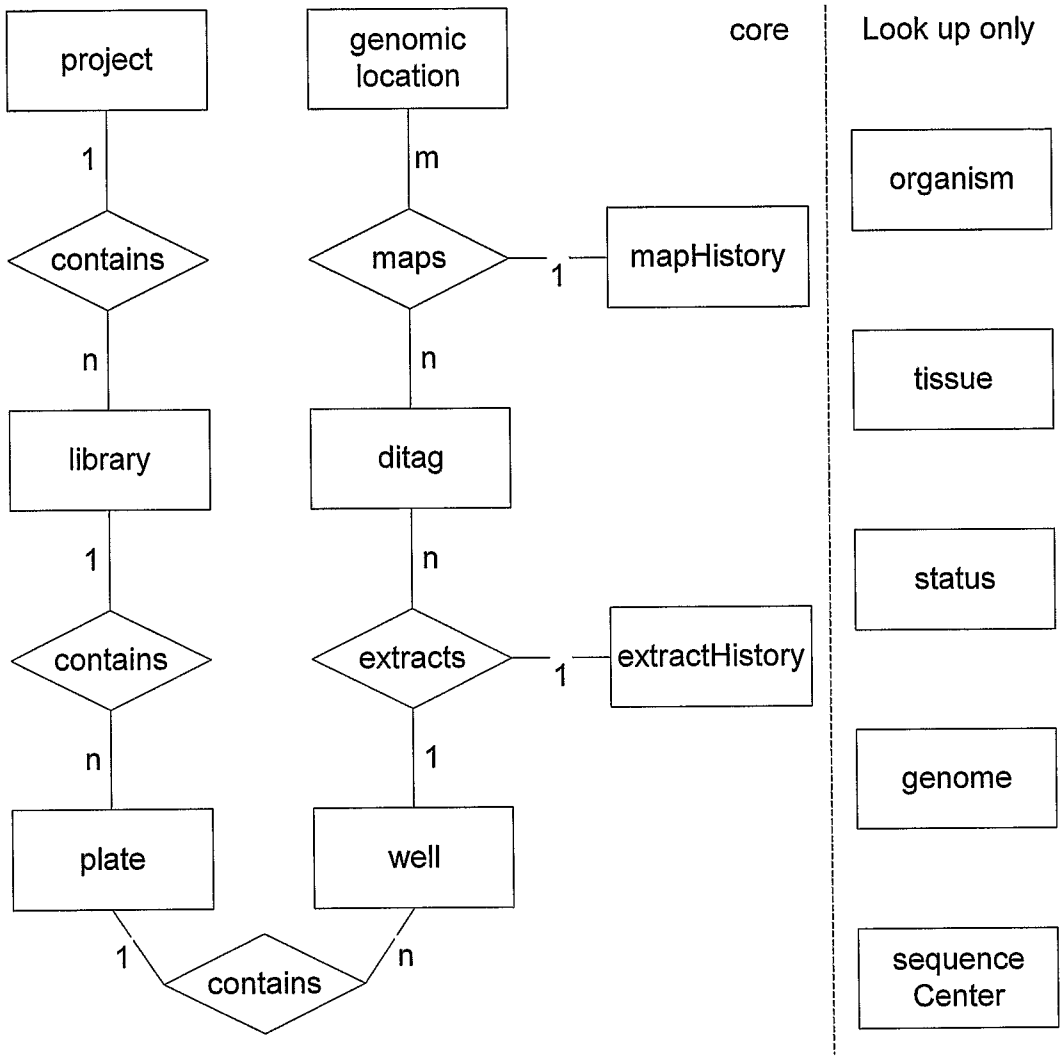

project
genomic location
core
Look up only
1
m
contains
maps
1
mapHistory
organism
n
n
tissue
library
ditag
1
n
status
contains
extracts
1
extractHistory
n
1
genome
plate
well
1
contains
n
sequence Center

FIGURE 10

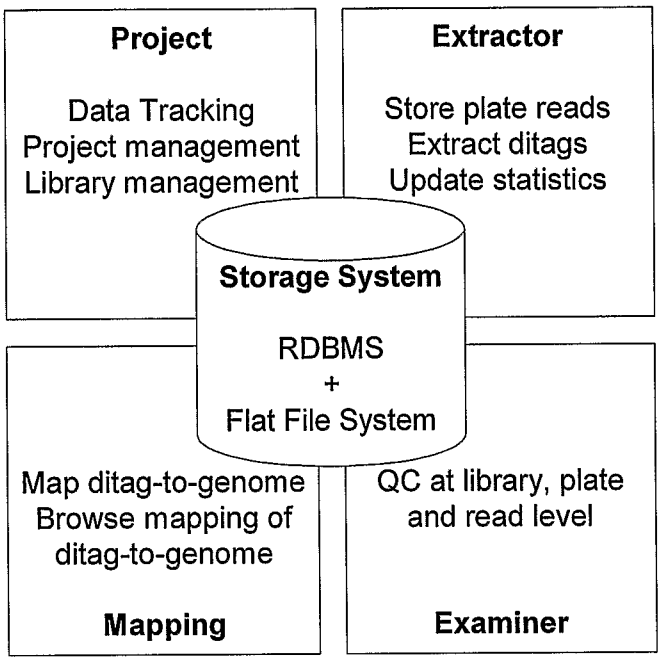

Project
Data Tracking
Project management
Library management
Extractor
Store plate reads
Extract ditags
Update statistics
Storage System
RDBMS
+
Flat File System
Map ditag-to-genome
Browse mapping of
ditag-to-genome
Mapping
QC at library, plate
and read level
Examiner

FIGURE 11

1. Project Module flow chart

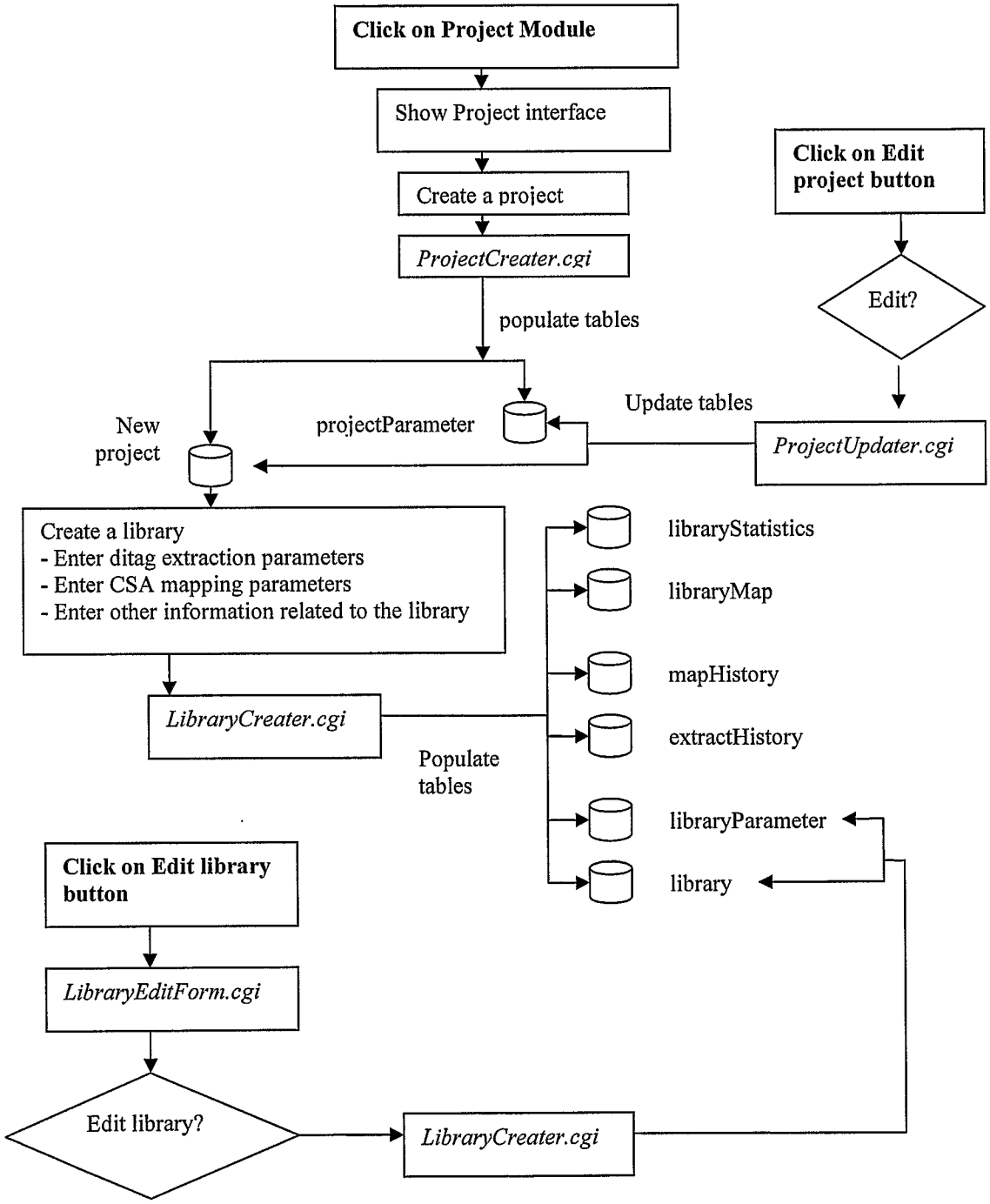


FIGURE 12

2. Extractor Module Flow Chart

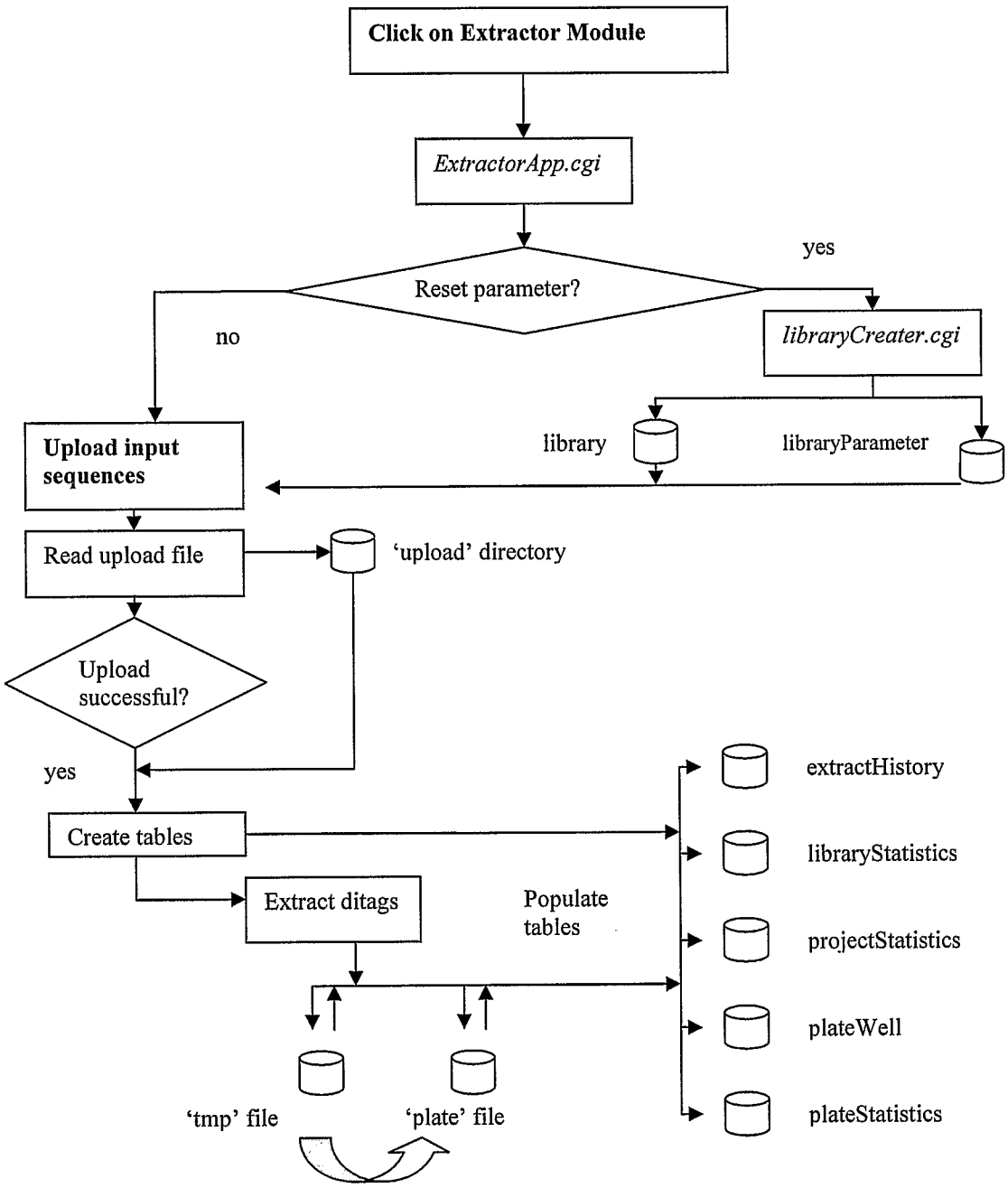


FIGURE 13

3. Examiner Module Flow Chart

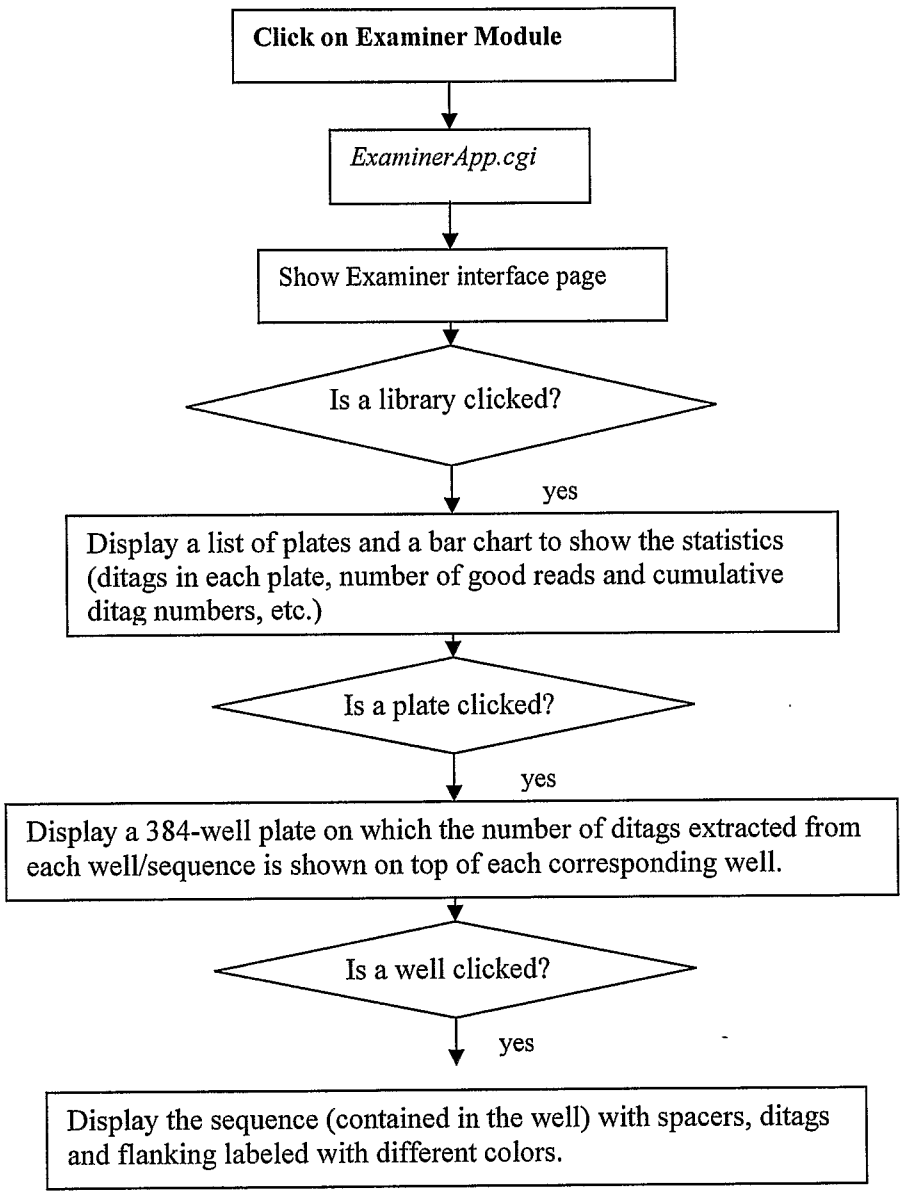


FIGURE 14

4. Mapping Module Flow Chart

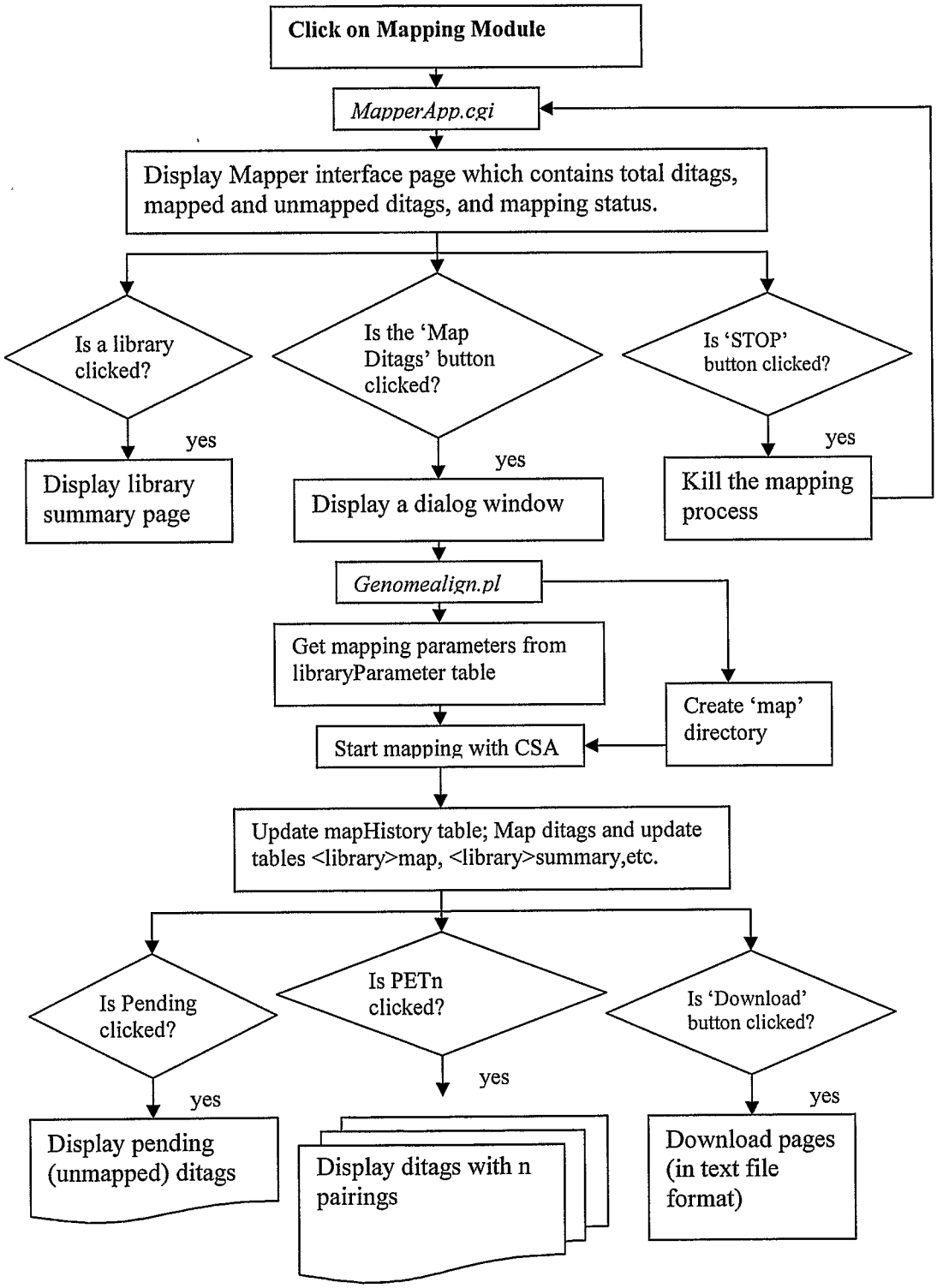


FIGURE 15

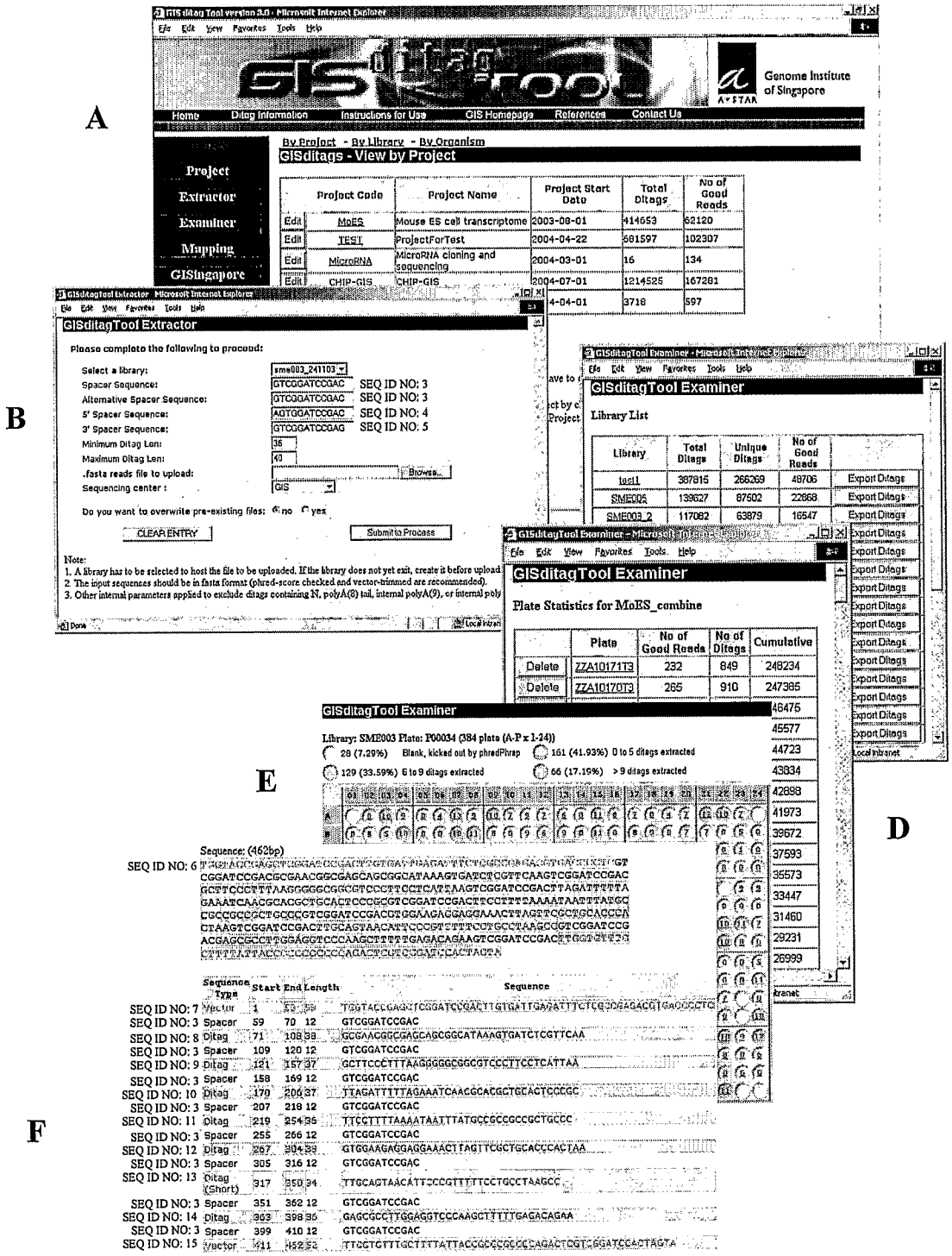
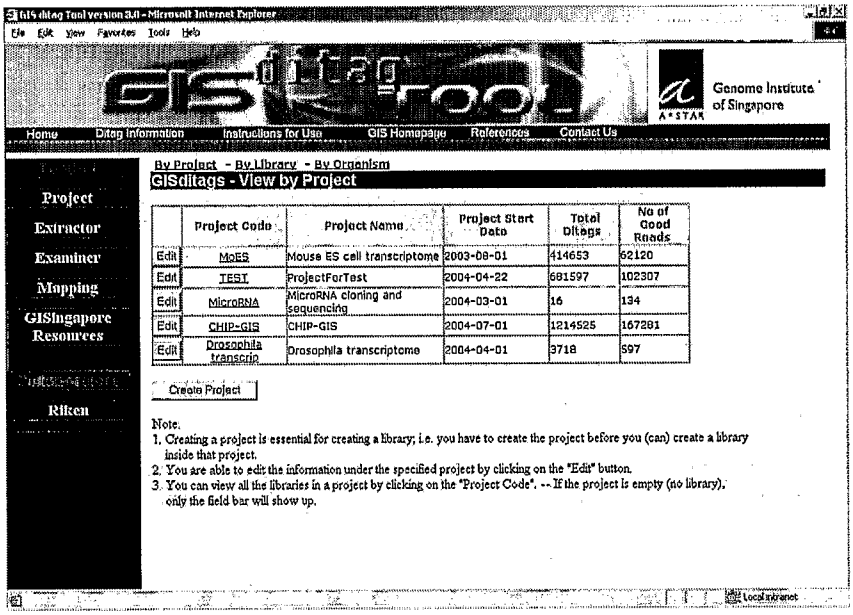

A
B
C
D
E
F
GISditagTool
Genome Institute of Singapore
Home
Ditag Information
Instructions for Use
GIS Homepage
References
Contact Us
Project
Extractor
Examiner
Mapping
GISingapore
By Project - By Library - By Organism
GISditags - View by Project
Project Code
Project Name
Project Start Date
Total Ditags
No of Good Reads
Edit MoES Mouse ES cell transcriptome 2003-08-01 414653 62120
Edit TEST ProjectForTest 2004-04-22 681597 102307
Edit MicroRNA MicroRNA cloning and sequencing 2004-03-01 16 134
Edit CHIP-GIS CHIP-GIS 2004-07-01 1214525 167281
4-04-01 3718 597
GISditagTool Extractor
Please complete the following to proceed:
Select a library:
Spacer Sequence: GTCGGATCCGAC SEQ ID NO: 3
Alternative Spacer Sequence: GTCGGATCCGAC SEQ ID NO: 3
5' Spacer Sequence: AGTGGATCCGAC SEQ ID NO: 4
3' Spacer Sequence: GTCGGATCCGAG SEQ ID NO: 5
Minimum Ditag Len:
Maximum Ditag Len:
.fasta reads file to upload:
Browse...
Sequencing center :
Do you want to overwrite pre-existing files: no yes
CLEAR ENTRY
Submit to Process
Note:
1. A library has to be selected to host the file to be uploaded. If the library does not yet exist, create it before upload.
2. The input sequences should be in fasta format (phred-score checked and vector-trimmed are recommended).
3. Other internal parameters applied to exclude ditags containing N, polyA(8) tail, internal polyA(9), or internal poly
GISditagTool Examiner
Library List
Library
Total Ditags
Unique Ditags
No of Good Reads
test1 387815 266269 48706 Export Ditags
SME005 139627 67502 22668 Export Ditags
SME003_2 117082 63879 16547 Export Ditags
GISditagTool Examiner
Plate Statistics for MoES_combine
Plate
No of Good Reads
No of Ditags
Cumulative
Delete ZZA10171T3 232 849 248234
Delete ZZA10170T3 265 910 247385
GISditagTool Examiner
Library: SME003 Plate: P00034 (384 plats (A-P x 1-24))
28 (7.29%) Blank, kicked out by phredPhrap
161 (41.93%) 0 to 5 ditags extracted
129 (33.59%) 6 to 9 ditags extracted
66 (17.19%) > 9 ditags extracted
Sequence: (462bp)
SEQ ID NO: 6
Sequence Type
Start
End
Length
Sequence
SEQ ID NO: 7 Vector 1 58 58 TGGTACCGAGCTCGGATCCGACTTGTGAATTGAATTTCTCGCGGAGACGTGACGCCTG
SEQ ID NO: 3 Spacer 59 70 12 GTCGGATCCGAC
SEQ ID NO: 8 Ditag 71 108 38 GCGAACGGCGAGCAGCGGCATAAAGTGATCTCGTTCAA
SEQ ID NO: 3 Spacer 109 120 12 GTCGGATCCGAC
SEQ ID NO: 9 Ditag 121 157 37 GCTTCCCTTTAAGGGGGCGGCGTCCCTTCCTCATTAA
SEQ ID NO: 3 Spacer 158 169 12 GTCGGATCCGAC
SEQ ID NO: 10 Ditag 170 206 37 TTAGATTTTTAGAAATCAACGCACGCTGCACTCCGCC
SEQ ID NO: 3 Spacer 207 218 12 GTCGGATCCGAC
SEQ ID NO: 11 Ditag 219 254 36 TTCGTTTTAAAATAATTTATGCCGCCGCCGCTGCCC
SEQ ID NO: 3 Spacer 255 266 12 GTCGGATCCGAC
SEQ ID NO: 12 Ditag 267 304 38 GTGGAAGAGGAGGAAACTTAGTTCGCTGCACCCACTAA
SEQ ID NO: 3 Spacer 305 316 12 GTCGGATCCGAC
SEQ ID NO: 13 Ditag (Short) 317 350 34 TTGCAGTAACATTCCGTTTTTCCTGCCTAAGCC
SEQ ID NO: 3 Spacer 351 362 12 GTCGGATCCGAC
SEQ ID NO: 14 Ditag 363 398 36 GAGCGCCTTGGAGGTCCCAAGCTTTTTGAGACAGAA
SEQ ID NO: 3 Spacer 399 410 12 GTCGGATCCGAC
SEQ ID NO: 15 Vector 411 462 52 TTCGTGTTTGCTTTTATTACCGCCCGCCCAGACTCGTCGGATCCACTAGTA

FIGURE 16

A


GIS ditag Tool version 3.0 - Microsoft Internet Explorer
File Edit View Favorites Tools Help
GIS ditag TOOL
Genome Institute of Singapore
A*STAR
Home
Ditag Information
Instructions for Use
GIS Homepage
References
Contact Us
Project
Extractor
Examiner
Mapping
GISingapore Resources
Riken
By Project - By Library - By Organism
GISditags - View by Project
Project Code
Project Name
Project Start Date
Total Ditags
No of Good Reads
Edit
MoES
Mouse ES cell transcriptome
2003-08-01
414653
62120
Edit
TEST
ProjectForTest
2004-04-22
681597
102307
Edit
MicroRNA
MicroRNA cloning and sequencing
2004-03-01
16
134
Edit
CHIP-GIS
CHIP-GIS
2004-07-01
1214525
167281
Edit
Drosophila transcrip
Drosophila transcriptome
2004-04-01
3718
597
Create Project
Note:
1. Creating a project is essential for creating a library; i.e. you have to create the project before you (can) create a library inside that project.
2. You are able to edit the information under the specified project by clicking on the "Edit" button.
3. You can view all the libraries in a project by clicking on the "Project Code". -- If the project is empty (no library), only the field bar will show up.
Local intranet

B

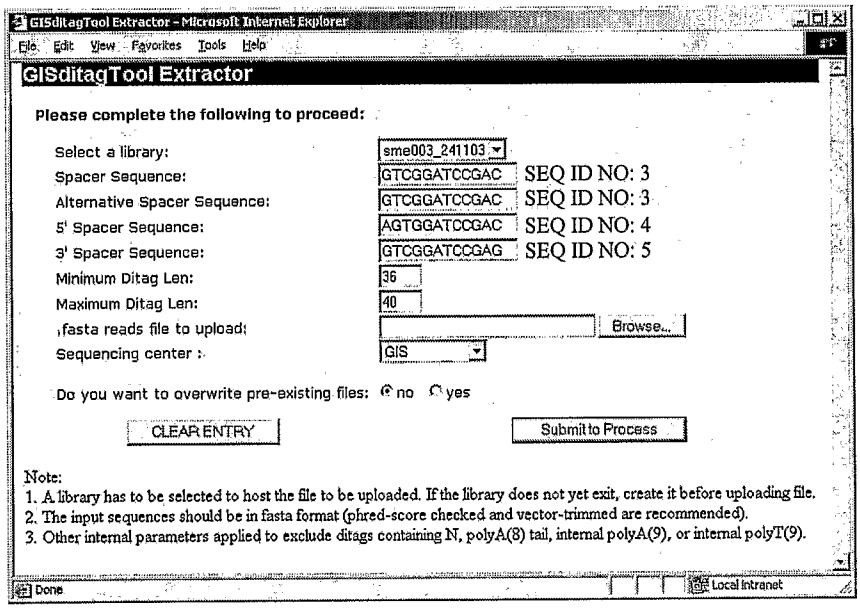
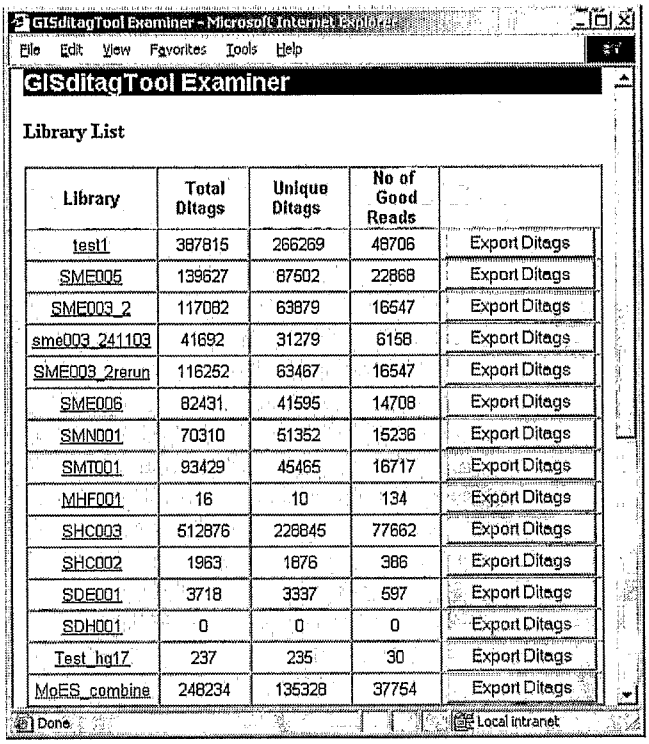

GISditagTool Extractor - Microsoft Internet Explorer
File Edit View Favorites Tools Help
GISditagTool Extractor
Please complete the following to proceed:
Select a library:
sme003_241103
Spacer Sequence:
GTCGGATCCGAC
SEQ ID NO: 3
Alternative Spacer Sequence:
GTCGGATCCGAC
SEQ ID NO: 3
5' Spacer Sequence:
AGTGGATCCGAC
SEQ ID NO: 4
3' Spacer Sequence:
GTCGGATCCGAG
SEQ ID NO: 5
Minimum Ditag Len:
36
Maximum Ditag Len:
40
,fasta reads file to upload:
Browse...
Sequencing center :
GIS
Do you want to overwrite pre-existing files: no yes
CLEAR ENTRY
Submit to Process
Note:
1. A library has to be selected to host the file to be uploaded. If the library does not yet exit, create it before uploading file.
2. The input sequences should be in fasta format (phred-score checked and vector-trimmed are recommended).
3. Other internal parameters applied to exclude ditags containing N, polyA(8) tail, internal polyA(9), or internal polyT(9).
Done
Local intranet

FIGURE 17

C



Library	Total Ditags	Unique Ditags	No of Good Reads	
test1	387815	266269	48706	Export Ditags
SME005	139627	87502	22868	Export Ditags
SME003_2	117082	63879	16547	Export Ditags
sme003_241103	41692	31279	6158	Export Ditags
SME003_2rerun	116252	63467	16547	Export Ditags
SME006	82431	41595	14708	Export Ditags
SMN001	70310	51352	15236	Export Ditags
SMT001	93429	45465	16717	Export Ditags
MHF001	16	10	134	Export Ditags
SHC003	512876	226845	77662	Export Ditags
SHC002	1963	1876	386	Export Ditags
SDE001	3718	3337	597	Export Ditags
SDH001	0	0	0	Export Ditags
Test_hg17	237	235	30	Export Ditags
MoES_combine	248234	135328	37754	Export Ditags

D

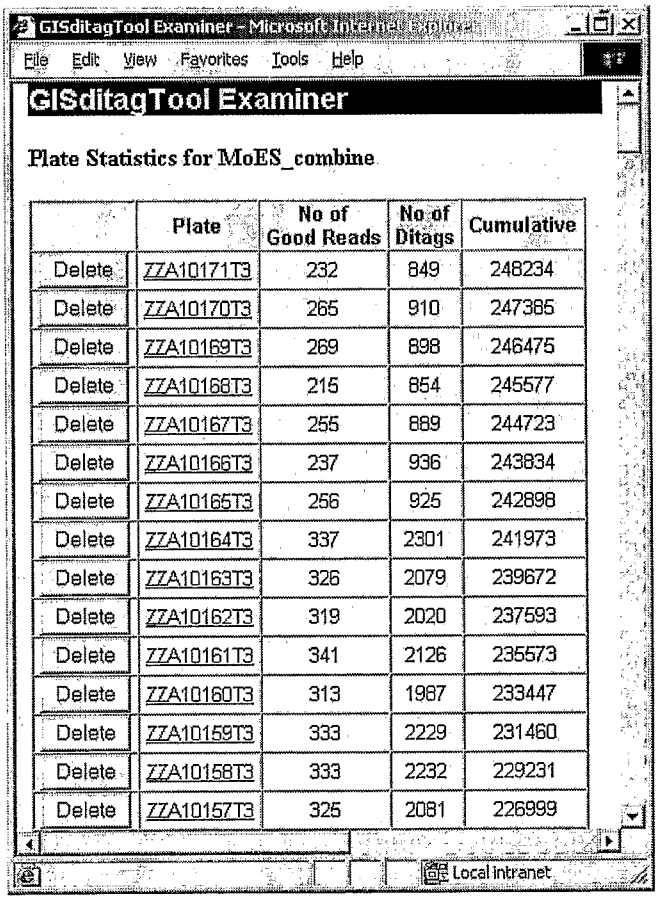


	Plate	No of Good Reads	No of Ditags	Cumulative
Delete	ZZA10171T3	232	849	248234
Delete	ZZA10170T3	265	910	247385
Delete	ZZA10169T3	269	898	246475
Delete	ZZA10168T3	215	854	245577
Delete	ZZA10167T3	255	889	244723
Delete	ZZA10166T3	237	936	243834
Delete	ZZA10165T3	256	925	242898
Delete	ZZA10164T3	337	2301	241973
Delete	ZZA10163T3	326	2079	239672
Delete	ZZA10162T3	319	2020	237593
Delete	ZZA10161T3	341	2126	235573
Delete	ZZA10160T3	313	1987	233447
Delete	ZZA10159T3	333	2229	231460
Delete	ZZA10158T3	333	2232	229231
Delete	ZZA10157T3	325	2081	226999

FIGURE 18

E

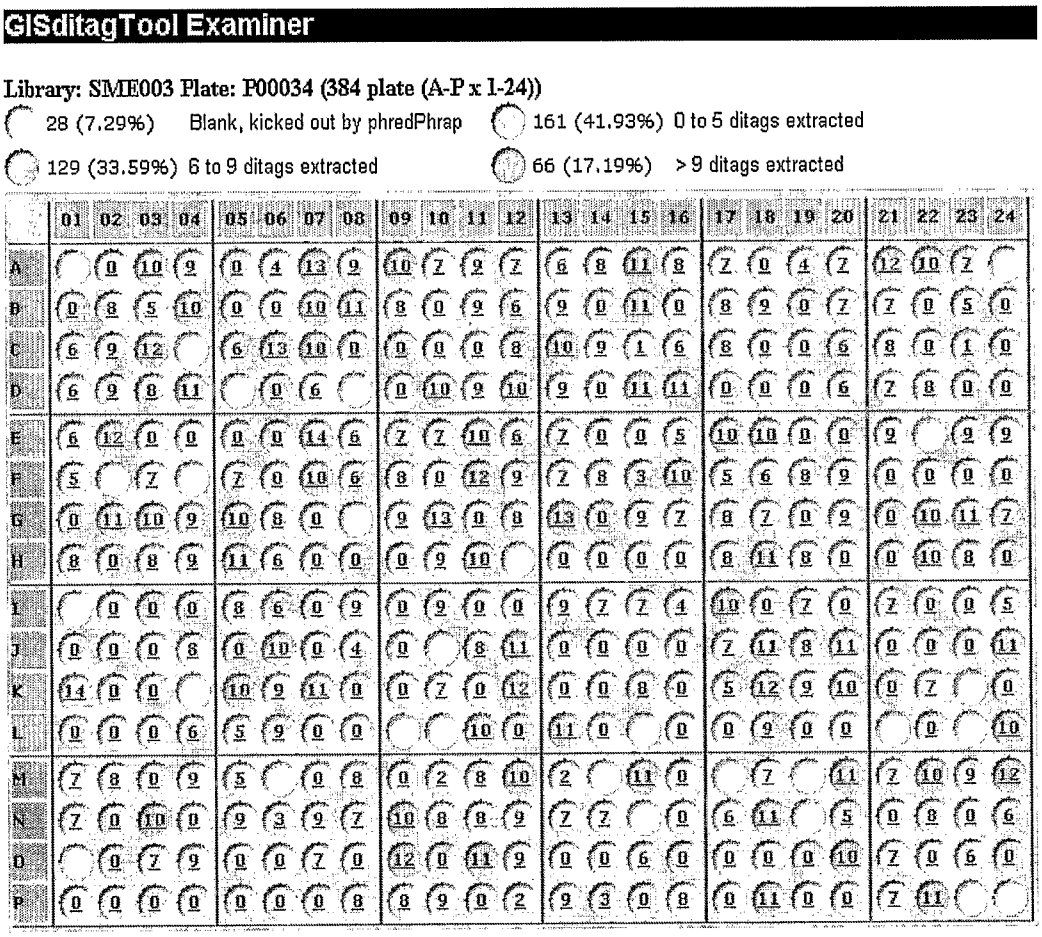

GISditagTool Examiner
Library: SME003 Plate: P00034 (384 plate (A-P x 1-24))
28 (7.29%) Blank, kicked out by phredPhrap
161 (41.93%) 0 to 5 ditags extracted
129 (33.59%) 6 to 9 ditags extracted
66 (17.19%) >9 ditags extracted
01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
A 0 10 9 0 4 13 9 10 7 9 7 6 8 11 8 7 0 4 7 12 10 7
B 0 8 5 10 0 0 10 11 8 0 9 6 9 0 11 0 8 9 0 7 7 0 5 0
C 6 9 12 6 13 10 0 0 0 0 8 10 9 1 6 8 0 0 6 8 0 1 0
D 6 9 8 11 0 6 0 10 9 10 9 0 11 11 0 0 0 6 7 8 0 0
E 6 12 0 0 0 0 14 6 7 7 10 6 7 0 0 5 10 10 0 0 9 9 9
F 5 7 7 0 10 6 8 0 12 9 7 8 3 10 5 6 8 9 0 0 0 0
G 0 11 10 9 10 8 0 9 13 0 8 13 0 9 7 8 7 0 9 0 10 11 7
H 8 0 8 9 11 6 0 0 0 9 10 0 0 0 0 8 11 8 0 0 10 8 0
I 0 0 0 8 6 0 9 0 9 0 0 9 7 7 4 10 0 7 0 7 0 0 5
J 0 0 0 8 0 10 0 4 0 8 11 0 0 0 0 7 11 8 11 0 0 0 11
K 14 0 0 10 9 11 0 0 7 0 12 0 0 8 0 5 12 9 10 0 7 0
L 0 0 0 6 5 9 0 0 10 0 11 0 0 9 0 0 0 0 10
M 7 8 0 9 5 0 8 0 2 8 10 2 11 0 7 11 7 10 9 12
N 7 0 10 0 9 3 9 7 10 8 8 9 7 7 0 6 11 5 0 8 0 6
O 0 7 9 0 0 7 0 12 0 11 9 0 0 6 0 0 0 0 10 7 0 6 0
P 0 0 0 0 0 0 0 8 8 9 0 2 9 3 0 8 0 11 0 0 7 11

FIGURE 19

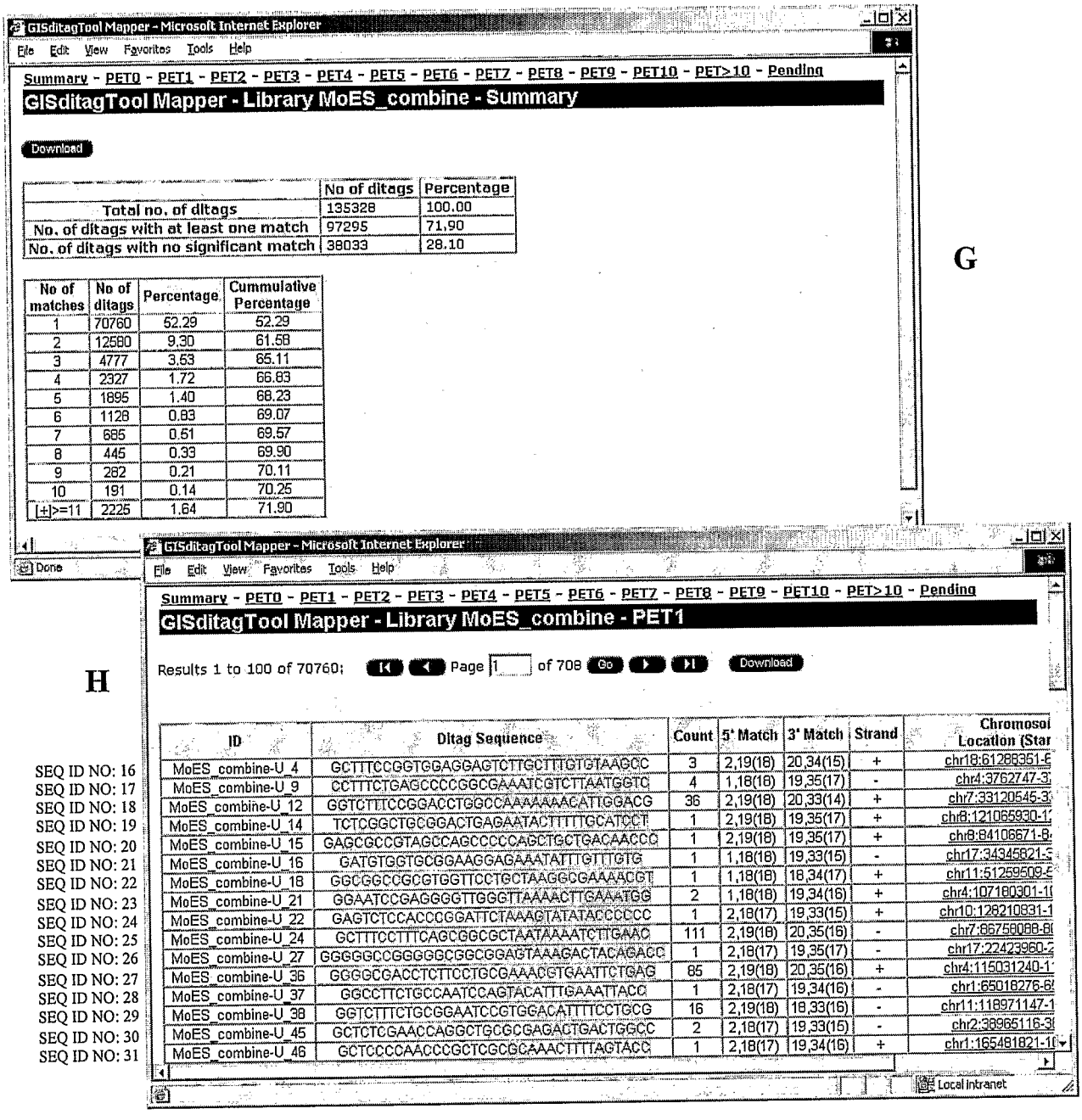
Sequence: (462bp)

TGGTACCGAGCTCGGATCCGACTTGTGATTGAGATTTCTCGCCGAGACGTGACCCCTCGT SEQ ID NO: 6
CGGATCCGACGCGAACGGCGAGCAGCGGCATAAAGTGATCTCGTTCAAGTCGGATCCGAC
GCTTCCCTTTAAGGGGGCGGCGTCCCTTCCTCATTAAGTCGGATCCGACTTAGATTTTTA
GAAATCAACGCACGCTGCACTCCCGCGTCGGATCCGACTTCCTTTTAAAATAATTTATGC
CGCCGCCGCTGCCCGTCGGATCCGACGTGGAAGAGGAGGAAACTTAGTTCGCTGCACCCA
CTAAGTCGGATCCGACTTGCAGTAACATTCCCGTTTTTCCTGCCTAAGCCGTCGGATCCG
ACGAGCGCCTTGGAGGTCCCAAGCTTTTTGAGACAGAAGTCGGATCCGACTTGGTGTTTG
CTTTTATTACCGCGCGCCCCAGACTCGTCGGATCCACTAGTA

F

Sequence Type	Start	End	Length	Sequence	
Vector	1	58	58	TGGTACCGAGCTCGGATCCGACTTGTGATTGAGATTTCTCGCCGAGACGTGACCCCTC	SEQ ID NO: 7
Spacer	59	70	12	GTCGGATCCGAC	SEQ ID NO: 3
Ditag	71	108	38	GCGAACGGCGAGCAGCGGCATAAAGTGATCTCGTTCAA	SEQ ID NO: 8
Spacer	109	120	12	GTCGGATCCGAC	SEQ ID NO: 3
Ditag	121	157	37	GCTTCCCTTTAAGGGGGCGGCGTCCCTTCCTCATTAA	SEQ ID NO: 9
Spacer	158	169	12	GTCGGATCCGAC	SEQ ID NO: 3
Ditag	170	206	37	TTAGATTTTTAGAAATCAACGCACGCTGCACTCCCGC	SEQ ID NO: 10
Spacer	207	218	12	GTCGGATCCGAC	SEQ ID NO: 3
Ditag	219	254	36	TTCCTTTTAAAATAATTTATGCCGCCGCCGCTGCCC	SEQ ID NO: 11
Spacer	255	266	12	GTCGGATCCGAC	SEQ ID NO: 3
Ditag	267	304	38	GTGGAAGAGGAGGAAACTTAGTTCGCTGCACCCACTAA	SEQ ID NO: 12
Spacer	305	316	12	GTCGGATCCGAC	SEQ ID NO: 3
Ditag (Short)	317	350	34	TTGCAGTAACATTCCCGTTTTTCCTGCCTAAGCC	SEQ ID NO: 13
Spacer	351	362	12	GTCGGATCCGAC	SEQ ID NO: 3
Ditag	363	398	36	GAGCGCCTTGGAGGTCCCAAGCTTTTTGAGACAGAA	SEQ ID NO: 14
Spacer	399	410	12	GTCGGATCCGAC	SEQ ID NO: 3
Vector	411	462	52	TTGGTGTTTGCTTTTATTACCGCGCGCCCCAGACTCGTCGGATCCACTAGTA	SEQ ID NO: 15

FIGURE 20


GISditagTool Mapper - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Summary - PET0 - PET1 - PET2 - PET3 - PET4 - PET5 - PET6 - PET7 - PET8 - PET9 - PET10 - PET>10 - Pending
GISditagTool Mapper - Library MoES_combine - Summary
Download
No of ditags Percentage
Total no. of ditags 135328 100.00
No. of ditags with at least one match 97295 71.90
No. of ditags with no significant match 38033 28.10
No of matches No of ditags Percentage Cummulative Percentage
1 70760 52.29 52.29
2 12580 9.30 61.58
3 4777 3.53 65.11
4 2327 1.72 66.83
5 1895 1.40 68.23
6 1128 0.83 69.07
7 685 0.51 69.57
8 445 0.33 69.90
9 282 0.21 70.11
10 191 0.14 70.25
[+]>=11 2225 1.64 71.90
G
Done
GISditagTool Mapper - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Summary - PET0 - PET1 - PET2 - PET3 - PET4 - PET5 - PET6 - PET7 - PET8 - PET9 - PET10 - PET>10 - Pending
GISditagTool Mapper - Library MoES_combine - PET1
Results 1 to 100 of 70760: Page 1 of 708 Go Download
H
ID Ditag Sequence Count 5' Match 3' Match Strand Chromosol Location (Star
MoES_combine-U_4 GCTTTCCGGTGGAGGAGTCTTGCTTTGTGTAAGCC 3 2,19(18) 20,34(15) + chr18:61286351-6
MoES_combine-U_9 CCTTTCTGAGCCCCGGCGAAATCGTCTTAATGGTC 4 1,18(18) 19,35(17) - chr4:3762747-3
MoES_combine-U_12 GGTCTTTCCGGACCTGGCCAAAAAAACATTGGACG 36 2,19(18) 20,33(14) + chr7:33120545-3
MoES_combine-U_14 TCTCGGCTGCGGACTGAGAATACTTTTTGCATCCT 1 2,19(18) 19,35(17) + chr8:121065930-1
MoES_combine-U_15 GAGCGCCGTAGCCAGCCCCCAGCTGCTGACAACCC 1 2,19(18) 19,35(17) + chr8:84106671-8
MoES_combine-U_16 GATGTGGTGCGGAAGGAGAAATATTTGTTTGTG 1 1,18(18) 19,33(15) - chr17:34345821-3
MoES_combine-U_18 GGCGGCCGCGTGGTTCCTGCTAAGGCGAAAACGT 1 1,18(18) 18,34(17) + chr11:51259509-5
MoES_combine-U_21 GGAATCCGAGGGGTTGGGTTAAAACTTGAAATGG 2 1,18(18) 19,34(16) + chr4:107180301-1
MoES_combine-U_22 GAGTCTCCACCCGGATTCTAAAGTATATACCCCCC 1 2,18(17) 19,33(15) + chr10:128210831-1
MoES_combine-U_24 GCTTTCCTTTCAGCGGCGCTAATAAAATCTTGAAC 111 2,19(18) 20,35(16) - chr7:66758088-8
MoES_combine-U_27 GGGGGCCGGGGGCGGCGGAGTAAAGACTACAGACC 1 2,18(17) 19,35(17) - chr17:22423960-2
MoES_combine-U_36 GGGGCGACCTCTTCCTGCGAAACGTGAATTCTGAG 85 2,19(18) 20,35(16) + chr4:115031240-1
MoES_combine-U_37 GGCCTTCTGCCAATCCAGTACATTTGAAATTACC 1 2,18(17) 19,34(16) - chr1:65018276-6
MoES_combine-U_38 GGTCTTTCTGCGGAATCCGTGGACATTTTCCTGCG 16 2,19(18) 18,33(16) - chr11:118971147-1
MoES_combine-U_45 GCTCTCGAACCAGGCTGCGCGAGACTGACTGGCC 2 2,18(17) 19,33(15) - chr2:38965116-3
MoES_combine-U_46 GCTCCCCAACCCGCTCGCGCAAACTTTAGTACC 1 2,18(17) 19,34(16) + chr1:165481821-1
Local intranet
SEQ ID NO: 16
SEQ ID NO: 17
SEQ ID NO: 18
SEQ ID NO: 19
SEQ ID NO: 20
SEQ ID NO: 21
SEQ ID NO: 22
SEQ ID NO: 23
SEQ ID NO: 24
SEQ ID NO: 25
SEQ ID NO: 26
SEQ ID NO: 27
SEQ ID NO: 28
SEQ ID NO: 29
SEQ ID NO: 30
SEQ ID NO: 31

SEQUENCE LISTING

<110> Agency for Science, Technology and Research

<120> Method of Processing and/or Genome Mapping of Ditag Sequences

<130> FP3064

<150> US 11/151,591
<151> 2005-06-14

<160> 31

<170> PatentIn version 3.3

<210> 1
<211> 36
<212> DNA
<213> Mus musculus

<400> 1
gcgctagagg cggcggcact acaagtttaa tatgaa 36

<210> 2
<211> 37
<212> DNA
<213> Mus musculus

<400> 2
gccttttcct gtgggagcaa acatgtggtg acaagaa 37

<210> 3
<211> 12
<212> DNA
<213> Mus musculus

<400> 3
gtcggatccg ac 12

<210> 4
<211> 12
<212> DNA
<213> Mus musculus

<400> 4
agtggatccg ac 12

<210> 5
<211> 12
<212> DNA
<213> Mus musculus

<400> 5
gtcggatccg ag 12

<210> 6

<211> 462
<212> DNA
<213> Mus musculus

<400> 6
tggtaccgag ctcggatccg acttgtgatt gagatttctc gccgagacgt gacccctcgt 60

cggatccgac gcgaacggcg agcagcggca taaagtgatc tcgttcaagt cggatccgac 120

gcttcccttt aaggggggcgg cgtcccttcc tcattaagtc ggatccgact tagattttta 180

gaaatcaacg cacgctgcac tcccgcgtcg gatccgactt ccttttaaaa taatttatgc 240

cgccgccgct gcccgtcgga tccgacgtgg aagaggagga aacttagttc gctgcaccca 300

ctaagtcgga tccgacttgc agtaacattc ccgtttttcc tgcctaagcc gtcggatccg 360

acgagcgcct tggaggtccc aagctttttg agacagaagt cggatccgac ttggtgtttg 420

cttttattac cgcgcgcccc agactcgtcg gatccactag ta 462

<210> 7
<211> 58
<212> DNA
<213> Mus musculus

<400> 7
tggtaccgag ctcggatccg acttgtgatt gagatttctc gccgagacgt gacccctc 58

<210> 8
<211> 38
<212> DNA
<213> Mus musculus

<400> 8
gcgaacggcg agcagcggca taaagtgatc tcgttcaa 38

<210> 9
<211> 37
<212> DNA
<213> Mus musculus

<400> 9
gcttcccttt aaggggcgg cgtcccttcc tcattaa 37

<210> 10
<211> 37
<212> DNA
<213> Mus musculus

<400> 10
ttagattttt agaaatcaac gcacgctgca ctcccgc 37

<210> 11
<211> 36
<212> DNA

<213> Mus musculus

<400> 11
ttccttttaa aataatttat gccgccgccg ctgccc 36

<210> 12
<211> 38
<212> DNA
<213> Mus musculus

<400> 12
gtggaagagg aggaaactta gttcgctgca cccactaa 38

<210> 13
<211> 34
<212> DNA
<213> Mus musculus

<400> 13
ttgcagtaac attcccgttt ttcctgccta agcc 34

<210> 14
<211> 36
<212> DNA
<213> Mus musculus

<400> 14
gagcgccttg gaggtcccaa gctttttgag acagaa 36

<210> 15
<211> 52
<212> DNA
<213> Mus musculus

<400> 15
ttggtgtttg cttttattac cgcgcgcccc agactcgtcg gatccactag ta 52

<210> 16
<211> 35
<212> DNA
<213> Mus musculus

<400> 16
gctttccggt ggaggagtct tgctttgtgt aagcc 35

<210> 17
<211> 35
<212> DNA
<213> Mus musculus

<400> 17
cctttctgag ccccggcgaa atcgtcttaa tggtc 35

<210> 18

<211> 35
<212> DNA
<213> Mus musculus

<400> 18
ggtctttccg gacctggcca aaaaaacatt ggacg 35

<210> 19
<211> 35
<212> DNA
<213> Mus musculus

<400> 19
tctcggctgc ggactgagaa tacttttttgc atcct 35

<210> 20
<211> 35
<212> DNA
<213> Mus musculus

<400> 20
gagcgccgta gccagccccc agctgctgac aaccc 35

<210> 21
<211> 33
<212> DNA
<213> Mus musculus

<400> 21
gatgtggtgc ggaaggagaa atatttgttt gtg 33

<210> 22
<211> 34
<212> DNA
<213> Mus musculus

<400> 22
ggcggccgcg tggttcctgc taaggcgaaa acgt 34

<210> 23
<211> 34
<212> DNA
<213> Mus musculus

<400> 23
ggaatccgag gggttgggtt aaaacttgaa atgg 34

<210> 24
<211> 35
<212> DNA
<213> Mus musculus

<400> 24
gagtctccac ccggattcta aagtatatac ccccc 35

<210> 25
<211> 35
<212> DNA
<213> Mus musculus

<400> 25
gctttccttt cagcggcgct aataaaatct tgaac 35

<210> 26
<211> 35
<212> DNA
<213> Mus musculus

<400> 26
gggggccggg ggcggcggag taaagactac agacc 35

<210> 27
<211> 35
<212> DNA
<213> Mus musculus

<400> 27
ggggcgacct cttcctgcga aacgtgaatt ctgag 35

<210> 28
<211> 34
<212> DNA
<213> Mus musculus

<400> 28
ggccttctgc caatccagta catttgaaat tacc 34

<210> 29
<211> 35
<212> DNA
<213> Mus musculus

<400> 29
ggtctttctg cggaatccgt ggacattttc ctgcg 35

<210> 30
<211> 34
<212> DNA
<213> Mus musculus

<400> 30
gctctcgaac caggctgcgc gagactgact ggcc 34

<210> 31
<211> 34
<212> DNA
<213> Mus musculus

<400> 31

gctccccaac ccgctcgcgc aaacttttag tacc 34