Genomics 70, 300–306 (2000)
doi:10.1006/geno.2000.6359, available online at http://www.idealibrary.com on IDEAL®

PANORAMA: An Integrated Web-Based Sequence Analysis Tool and Its Role in Gene Discovery

Alexander Pertsemlidis,*,†,‡ Ashwini Pande,§ Brady Miller,¶ Peter Schilling,‖ Ming Hui Wei,** Michael I. Lerman,** John D. Minna,†,††,‡‡ and Harold R. Garner*,†,‡,[1]

*Department of Biochemistry, †Department of Internal Medicine, ‡McDermott Center for Human Growth and Development, §Program in Biomedical Engineering, Southwestern Graduate School of Biomedical Sciences, ¶Southwestern Medical School, ‖Genome Science and Technology Center, ††Department of Pharmacology, and ‡‡Hamon Center for Therapeutic Oncology Research, University of Texas Southwestern Medical Center, Dallas, Texas 75390; and **Laboratory of Immunobiology, National Cancer Institute, Frederick Cancer Research and Development Center, Frederick, Maryland 21702

Received June 19, 2000; accepted August 16, 2000; published online November 15, 2000

As the exponential growth of DNA sequence information in databases continues, the task of converting this deposited information into knowledge becomes more dependent on integrative sequence analysis and visualization tools. PANORAMA is an Internet-accessible software package that performs a variety of informatics analyses on a given DNA sequence and returns a visual and interactive representation of the results. Its design is modular, so that further sequence analysis tools can be integrated with minimal effort. The utility of PANORAMA is demonstrated in the analysis of 650 kb of human genomic DNA from chromosome region 3p21.3, a region of potential tumor suppressor genes involved in lung cancer, breast cancer, and other forms of cancer. PANORAMA aided in the discovery of genes and alternate splice forms of known exons, in the demarcation of intron–exon boundaries, and in the identification of promoter regions and polymorphisms, all of which contributed to a better understanding of the region. PANORAMA is available on the World Wide Web at http://atlas.swmed.edu. © 2000 Academic Press

INTRODUCTION

Analysis of a DNA sequence using informatics involves the simultaneous identification and characterization of a number of attributes, including nucleotide content, homology or identity with known genes or expressed sequences, predicted structures, polymorphic markers, and anything else that brings a better understanding of the genes it contains and the regulation of their expression or the functions of the proteins that they encode. Identification of possible open reading frames and coding regions in the sequence allows prediction of the amino acid sequences that they generate and investigation of some aspects of the predicted protein products. Nucleotide content of the sequence can also give some indication about the DNA structure and the regulation of gene expression. Identification of any region of the sequence previously assigned a chromosomal location can aid genetic mapping, while identification of new polymorphic markers can be used to refine the location of a known locus or place a previously unknown sequence on the genetic recombination map. With the success of the Human Genome Project and private sequencing efforts, a usable draft of the total human genomic sequence is now available. The next step is the identification of all of the genes, a large part of which will depend on sequence analysis and annotation.

A wide variety of genomic sequence analysis and annotation tools are currently available. Some are independent tools that identify and analyze one particular feature of a sequence, such as BLAST (Altschul *et al.*, 1990), which finds homologies between sequences; GenScan (Burge and Karlin, 1997), which predicts cDNA and peptide sequences and the locations of introns and exons from genomic sequence; and POMPOUS (Fondon *et al.*, 1998), which finds polymorphic repeat regions and predicts primers. The assembly and comparison of the results of these individual applications are typically a tedious serial process that needs to be repeated with each sequence and database update, and requires switching between multiple different applications and Web sites. This underscores the need for a comprehensive tool that automatically executes various sequence analysis tools and gives combined results that are expressed graphically, are easy to interpret, and can be compared.

Other tools identify a number of features and include homology search against GenBank and gene predic-

[1] To whom correspondence should be addressed at Department of Biochemistry and Department of Internal Medicine, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390-8591. Telephone: (214) 648-1661. Fax: (214) 648-1445. E-mail: garner@utsw.swmed.edu.

0888-7543/00 $35.00

tion. None of these tools, however, includes identification of CpG islands, prediction of polymorphic regions, and prediction of DNA structures, all of which are implemented in PANORAMA. While some of the currently available tools are both comprehensive and available on the Internet, like the BCM Search Launcher (Smith *et al.*, 1996), GeneQuiz (Andrade *et al.*, 1999; Scharf *et al.*, 1994), and the Staden Package (Staden, 1996), most do not provide a combined graphical representation of the results, making the comparison of various results difficult. Genotator (Harris, 1997, 2000) provides features such as comparison of results from various gene prediction programs and identification of given patterns of nucleotides. Along with GAIA (Genome Annotation and Information Analysis) (Bailey *et al.*, 1998), it uses Java applets for the graphical output with features such as zooming and clicking on regions to obtain annotations. However, only Genotator is complete and available for downloading. Its disadvantage is that it requires local installation of BLAST, Genie, and other such sequence analysis applications to exploit its capabilities fully. In contrast, PANORAMA has all of the software and databases installed on a Web server, simply requiring the user to submit the genomic sequence to be analyzed.

MATERIALS AND METHODS

Computational tools and genomic DNA sequences. All programs used by PANORAMA to combine the various sequence analysis programs were written in C, Perl, or Java and run on an HP9000/898 K370 with four processors and 2 GB shared RAM. PANORAMA uses a parallel version of BLAST (Li *et al.*, 1997a) for sequence comparison. Databases, which are updated frequently, include all of GenBank split into an EST section and a non-EST section (which contains high-throughput genomic sequence data) and a set of repeat sequences (Jurka, 1998).

Program organization. PANORAMA is modular in design, allowing for expansion of its capabilities. In principle, any analysis application can be added as long as an output parser is provided. PANORAMA takes a query sequence and executes each of its analysis modules sequentially, compiling the results, and displaying them as text, static graphics, and through a clickable Java interface. A flowchart of the analysis performed by PANORAMA is shown in Fig. 1.

Identification and masking of mammalian repetitive elements and simple sequences. The query sequence is first masked for mammalian repetitive elements (REs) and simple sequences like Alu and LINE1, which otherwise would result in a large number of noninformative hits against GenBank sequences containing these repeat regions. These repeats are identified by using BLAST to compare the given sequence with a database of repeat sequences (Jurka, 1998). The repetitive regions found in the sequence are then masked by replacing individual bases with Ns. The resulting masked sequence is then compared against EST and non-EST sections of GenBank.

Comparison with EST and non-EST GenBank. A parallel version of BLAST (Altschul *et al.*, 1990, 1997; Li *et al.*, 1997a) is used to compare the query sequence with the EST and non-EST sequences. The BLAST results are parsed and filtered based on the identity and score of each hit. The thresholds for identity and score are user-definable parameters, with defaults of 90% and 140, respectively, which should eliminate spurious hits and reduce the clutter in the graphical output. The regions above these cutoffs are copied from the raw BLAST output to a regions database file and summarized in the features table, to be discussed later. Figure 2 is sample output that shows the hits in the EST GenBank from a query sequence of 28,241 bp of genomic DNA of a cosmid clone, LUCA12 (AC002481), from human chromosome region 3p21.3.

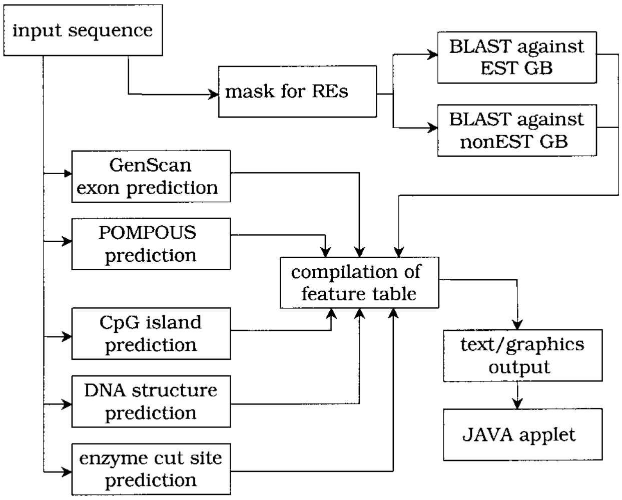


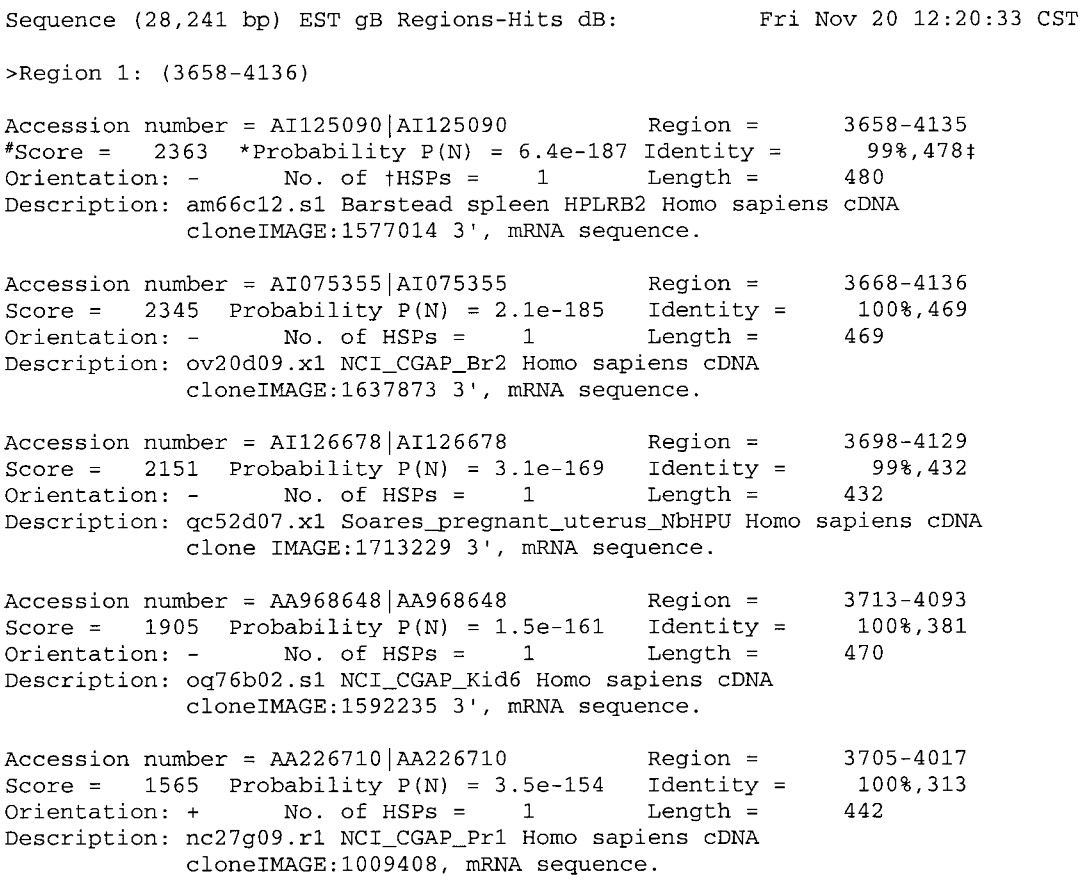
FIG. 1. PANORAMA flowchart. Each of the rectangular blocks indicates the various programs and scripts called by PANORAMA. RE stands for repetitive elements, and GB stands for GenBank.

Prediction of exons. GenScan (Burge and Karlin, 1997) is used for the prediction of introns, exons, initiation sites, polyadenylation sites, and promoter regions. GenScan is preferred over other gene prediction applications because of its speed, accuracy, and availability. GenScan was shown to have substantially higher accuracy than competing methods when tested on standardized sets of human and vertebrate genes, with 75 to 80% of exons identified exactly (Burge and Karlin, 1997; Murakami and Takagi, 1998). Displaying GenScan predictions together with the EST sequences aligned by homology searching can aid in the discovery of a more complete set of potential coding regions (exons) and therefore of previously unknown genes or alternatively spliced forms of known genes. Predicted polyadenylation sites and promoter regions are also shown as they provide supporting evidence that can help to define gene structure.

Prediction of polymorphic markers. Polymorphic markers in the form of simple sequence repeats are predicted by POMPOUS (Fondon *et al.*, 1998). Also predicted are several primers (Li *et al.*, 1997b) flanking these markers and melting temperatures that can be used to validate the potential polymorphisms via polymerase chain reaction. The location of the predicted polymorphic markers in relation to one another and to other specific features can be useful in precise mapping analyses such as loss of heterozygosity studies, which involve locating breakpoints relative to genes. Prior work has shown that POMPOUS is 67% accurate in predicting polymorphisms (Fondon *et al.*, 1998).

Detection of CpG islands. CpG islands are frequently located in regions regulating the expression of genes, such as promoter regions. They are commonly defined as regions of DNA of at least 200 bp in length that have a G+C content above 50% and a ratio of observed to expected CpGs (CG pairs) close to or above 0.6 (Gardiner-Garden and Frommer, 1987). PANORAMA identifies CpG islands using the above definition in an efficient algorithm that parses the input sequence only once (Larsen *et al.*, 1992). With PANORAMA, these CpG islands are displayed graphically and can be quickly correlated with other known genes and EST sequences present in GenBank. Identification of CpG islands also helps identify possible locations of new and previously undiscovered genes and sites to test for changes in methylation patterns of gene regulatory regions.

Prediction of DNA structures. The structure of DNA plays a significant role in gene regulation. In addition to the simple sequence repeats mentioned above, PANORAMA identifies three types of DNA structure: triplex DNA, tetraplex DNA, and Z-DNA. PANORAMA is



```
Sequence (28,241 bp) EST gB Regions-Hits dB:          Fri Nov 20 12:20:33 CST

>Region 1: (3658-4136)

Accession number = AI125090|AI125090          Region =       3658-4135
#Score =    2363  *Probability P(N) = 6.4e-187 Identity =       99%,478‡
Orientation: -       No. of †HSPs =     1       Length =       480
Description: am66c12.s1 Barstead spleen HPLRB2 Homo sapiens cDNA
             cloneIMAGE:1577014 3', mRNA sequence.

Accession number = AI075355|AI075355          Region =       3668-4136
Score =    2345  Probability P(N) = 2.1e-185   Identity =      100%,469
Orientation: -       No. of HSPs =     1       Length =       469
Description: ov20d09.x1 NCI_CGAP_Br2 Homo sapiens cDNA
             cloneIMAGE:1637873 3', mRNA sequence.

Accession number = AI126678|AI126678          Region =       3698-4129
Score =    2151  Probability P(N) = 3.1e-169   Identity =       99%,432
Orientation: -       No. of HSPs =     1       Length =       432
Description: qc52d07.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA
             clone IMAGE:1713229 3', mRNA sequence.

Accession number = AA968648|AA968648          Region =       3713-4093
Score =    1905  Probability P(N) = 1.5e-161   Identity =      100%,381
Orientation: -       No. of HSPs =     1       Length =       470
Description: oq76b02.s1 NCI_CGAP_Kid6 Homo sapiens cDNA
             cloneIMAGE:1592235 3', mRNA sequence.

Accession number = AA226710|AA226710          Region =       3705-4017
Score =    1565  Probability P(N) = 3.5e-154   Identity =      100%,313
Orientation: +       No. of HSPs =     1       Length =       442
Description: nc27g09.r1 NCI_CGAP_Pr1 Homo sapiens cDNA
             cloneIMAGE:1009408, mRNA sequence.
```

FIG. 2. Sample PANORAMA output showing the hits found in EST GenBank from the input of 28,241 bp of LUCA 12 (AC002481). #Score is a measure of the similarity of the two sequences being compared and is calculated by BLAST. †HSP indicates a high-scoring segment pair. **P*(*N*) is the smallest sum probability assigned to *N* HSPs identified. This is based on Karlin–Altschul statistics and indicates the probability that an alignment with the given score could be due simply to chance. ‡The value after the percentage identity gives the length of the HSP over which the identity holds.

not exhaustive in this prediction but tries to identify the most significant possibilities. To find triplex DNA, PANORAMA looks for the pattern (CT)$_n$ · (GA)$_n$, where *n* is at least 5 (Sinden, 1994). Intramolecular triplex DNA can form within a single homopurine–homopyrimidine duplex DNA region containing mirror repeat symmetry in supercoiled DNA. Intramolecular DNA triplexes may potentially affect the level of transcription and gene expression by means of proteins binding specifically to triplex DNA structures (Sinden, 1994; Ulrich *et al.*, 1992). To find tetraplex, or four-stranded, DNA, PANORAMA looks for the patterns (G$_4$T$_2$)$_n$ and (G$_4$T$_4$)$_n$, where *n* is at least 5 (Cherepanov *et al.*, 1997). Tetraplex DNA is formed by basepairing of one base with its complementary base as well as the complementary base of a second strand, such as occurs at telomeres (Sinden, 1994). To find Z-DNA, PANORAMA looks for the patterns (GC)$_n$ and (GT)$_n$, where *n* is at least 10. Z-DNA is a left-handed helix that can form in alternating purine–pyrimidine tracts, especially (GC)$_n$ and (GT)$_n$, under certain conditions, including high salt, the presence of certain divalent cations, or DNA supercoiling. The possible roles of Z-DNA may include controlling gene expression, influencing transcription, and influencing the possible positioning of nucleosomes (Sinden, 1994; Thomas *et al.*, 1990).

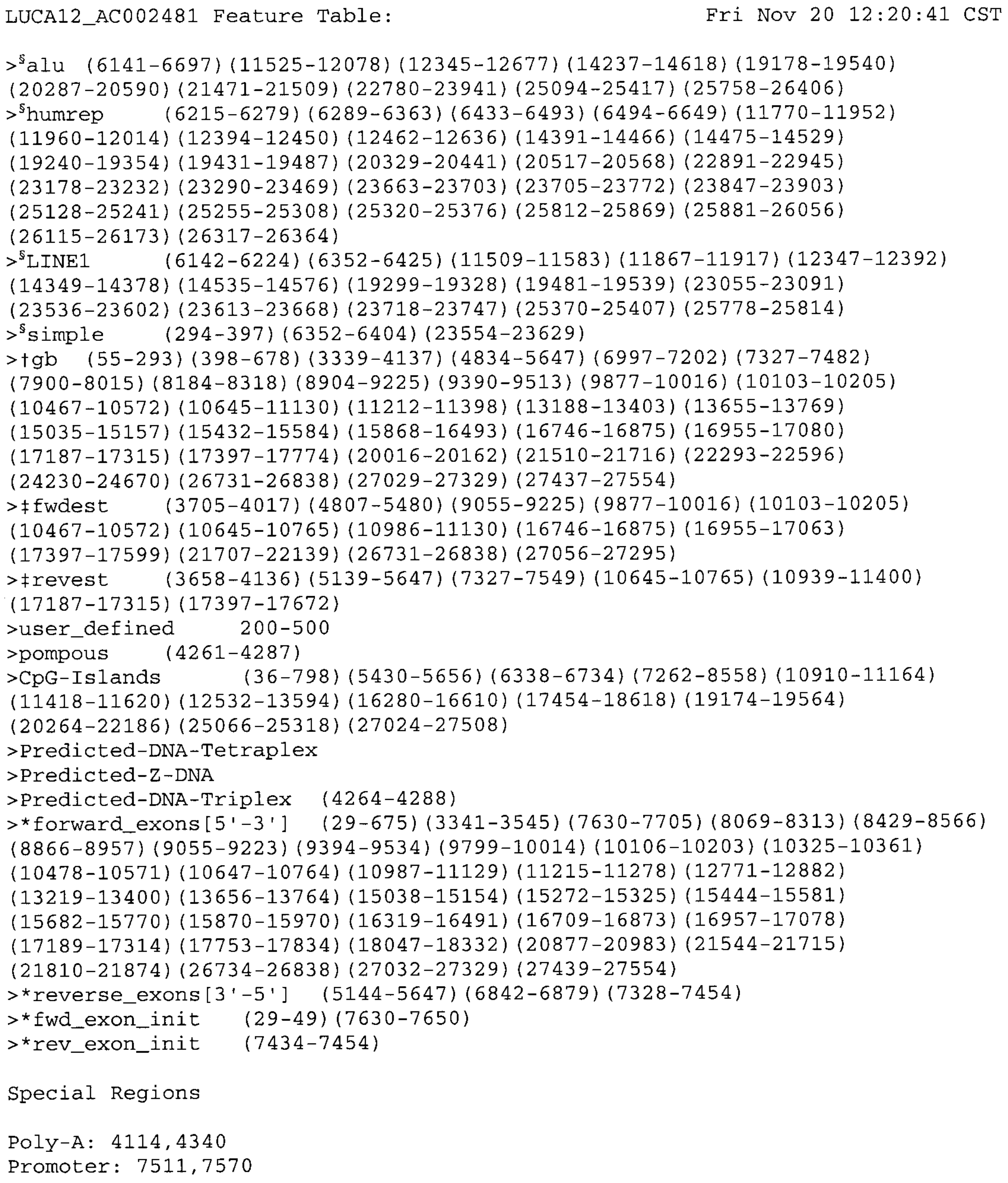
Identification of restriction enzyme cut sites. The identification of restriction enzyme cut sites can form a restriction map useful for molecular genetics experiments. The restriction fragment cut sites located by PANORAMA are *Bam*HI (G′GATC_C), *Eco*RI (G′AATT_C), *Not*I (GC′GGCC_GC), and *Sfi*I (GGCCN_NNN′NGGCC), where the prime symbol indicates the cut site and the underscore indicates the overhang. These restriction fragment cut sites were selected because they occur often in PAC/BAC multiple cloning sites and are used to verify sequence assemblies. The results are displayed in a separate file, but are not marked on the PANORAMA plot to preserve clarity.

Addition of user-defined regions. Addition of user-defined regions allows the user to indicate regions in the sequence to add to the features table and highlight in the graphical output. This option is typically used to add features that have been recently discovered but not published.

Filtering of given GenBank accession numbers. The user is allowed to list GenBank entries by accession numbers that should not be displayed on the graphical output because they would overlap or obscure other findings. This feature is useful when analyzing genomic sequences that are identical to or have significant homology to sequences already deposited in GenBank.

Preparation of the features table. The first step in generating an integrated output in PANORAMA is the creation of the features table. As each of the sequence analysis tools is applied to the given sequence, the results generated are extracted and summarized in the features table as lists of regions corresponding to each of the features identified (Fig. 3). This table is used to generate the graphical output and the precise location (in basepairs) of each feature.

Generation of the graphical output. A graphical plot of the features table is generated in Adobe Portable Document Format (PDF) and PostScript formats. The PostScript is constructed first, with a set of Perl scripts used to set the origin of the plot, calculate the sizes and positions of the bars for each feature to be displayed, and determine the resolution of the plot (the number of kilobases to be represented per inch, which depends on the sequence length). The PostScript is then converted to PDF format using the Ghostscript



```
LUCA12_AC002481 Feature Table:                      Fri Nov 20 12:20:41 CST

>§alu  (6141-6697)(11525-12078)(12345-12677)(14237-14618)(19178-19540)
(20287-20590)(21471-21509)(22780-23941)(25094-25417)(25758-26406)
>§humrep     (6215-6279)(6289-6363)(6433-6493)(6494-6649)(11770-11952)
(11960-12014)(12394-12450)(12462-12636)(14391-14466)(14475-14529)
(19240-19354)(19431-19487)(20329-20441)(20517-20568)(22891-22945)
(23178-23232)(23290-23469)(23663-23703)(23705-23772)(23847-23903)
(25128-25241)(25255-25308)(25320-25376)(25812-25869)(25881-26056)
(26115-26173)(26317-26364)
>§LINE1      (6142-6224)(6352-6425)(11509-11583)(11867-11917)(12347-12392)
(14349-14378)(14535-14576)(19299-19328)(19481-19539)(23055-23091)
(23536-23602)(23613-23668)(23718-23747)(25370-25407)(25778-25814)
>§simple     (294-397)(6352-6404)(23554-23629)
>†gb  (55-293)(398-678)(3339-4137)(4834-5647)(6997-7202)(7327-7482)
(7900-8015)(8184-8318)(8904-9225)(9390-9513)(9877-10016)(10103-10205)
(10467-10572)(10645-11130)(11212-11398)(13188-13403)(13655-13769)
(15035-15157)(15432-15584)(15868-16493)(16746-16875)(16955-17080)
(17187-17315)(17397-17774)(20016-20162)(21510-21716)(22293-22596)
(24230-24670)(26731-26838)(27029-27329)(27437-27554)
>‡fwdest     (3705-4017)(4807-5480)(9055-9225)(9877-10016)(10103-10205)
(10467-10572)(10645-10765)(10986-11130)(16746-16875)(16955-17063)
(17397-17599)(21707-22139)(26731-26838)(27056-27295)
>‡revest     (3658-4136)(5139-5647)(7327-7549)(10645-10765)(10939-11400)
(17187-17315)(17397-17672)
>user_defined      200-500
>pompous     (4261-4287)
>CpG-Islands       (36-798)(5430-5656)(6338-6734)(7262-8558)(10910-11164)
(11418-11620)(12532-13594)(16280-16610)(17454-18618)(19174-19564)
(20264-22186)(25066-25318)(27024-27508)
>Predicted-DNA-Tetraplex
>Predicted-Z-DNA
>Predicted-DNA-Triplex  (4264-4288)
>*forward_exons[5'-3']  (29-675)(3341-3545)(7630-7705)(8069-8313)(8429-8566)
(8866-8957)(9055-9223)(9394-9534)(9799-10014)(10106-10203)(10325-10361)
(10478-10571)(10647-10764)(10987-11129)(11215-11278)(12771-12882)
(13219-13400)(13656-13764)(15038-15154)(15272-15325)(15444-15581)
(15682-15770)(15870-15970)(16319-16491)(16709-16873)(16957-17078)
(17189-17314)(17753-17834)(18047-18332)(20877-20983)(21544-21715)
(21810-21874)(26734-26838)(27032-27329)(27439-27554)
>*reverse_exons[3'-5']  (5144-5647)(6842-6879)(7328-7454)
>*fwd_exon_init   (29-49)(7630-7650)
>*rev_exon_init   (7434-7454)

Special Regions

Poly-A: 4114,4340
Promoter: 7511,7570
```

FIG. 3. Sample features table generated by PANORAMA for the input of 28,241 bp of LUCA 12 (AC002481). §Simple repeats and human repetitive elements that are masked out of the sequence before the BLAST against GenBank. †gb: non-EST GenBank hits. ‡fwdest: 5′-3′ EST GenBank entry hits; revest: 3′-5′ EST GenBank entry hits. *GenScan predicted exons.

utility ps2pdf. The PDF file can be viewed using Adobe Acrobat Reader 3.01 or higher.

Figure 4 shows a sample PANORAMA plot for an input of 28,241 bp of genomic DNA sequence of cosmid LUCA12 (Accession No. AC002481). A black reference line indicates the sequence itself. Bars at different offsets from the reference line are used to represent the various characteristics. For example, one CpG island in the LUCA 12 (AC002481) sequence can be seen in Fig. 4 as a horizontal bar extending from base 7262 to base 8558. Hits against EST and non-EST GenBank sequences are shown, as are GenScan predicted sequences and human repetitive elements. The forward strands (5′-3′) of the GenScan predicted exons and the identified EST regions are above the reference line, while the reverse strands (3′-5′) are below the reference line. Small vertical lines below the reference line indicate poly(A) regions and promoters. A small vertical black line on the bars, shown in Fig. 4 at base positions 1, 7500, and 7600, is used to indicate the start of a predicted exon, to distinguish between adjacent exons in the same orientation.

Generation of interactive JAVA applet. All of the entries in the features table except the restriction enzyme cut sites, as noted above, are presented in an interactive JAVA applet that allows the user to edit, modify, and save the graphical output. The color bars are clickable and call up additional windows containing the BLAST results from which the bars were generated. The scale of the display, zoom, and spacing of the tick marks can be changed. There is a properties option, which allows the user to change the width and height of the displayed page, the number of nucleotides per line and per tick, and the height of the color bars. In addition, analysis tools can be removed, colors can be changed, user-defined regions can be added, the sequence name can be changed, and almost all actions can be undone (except the sequence name change).

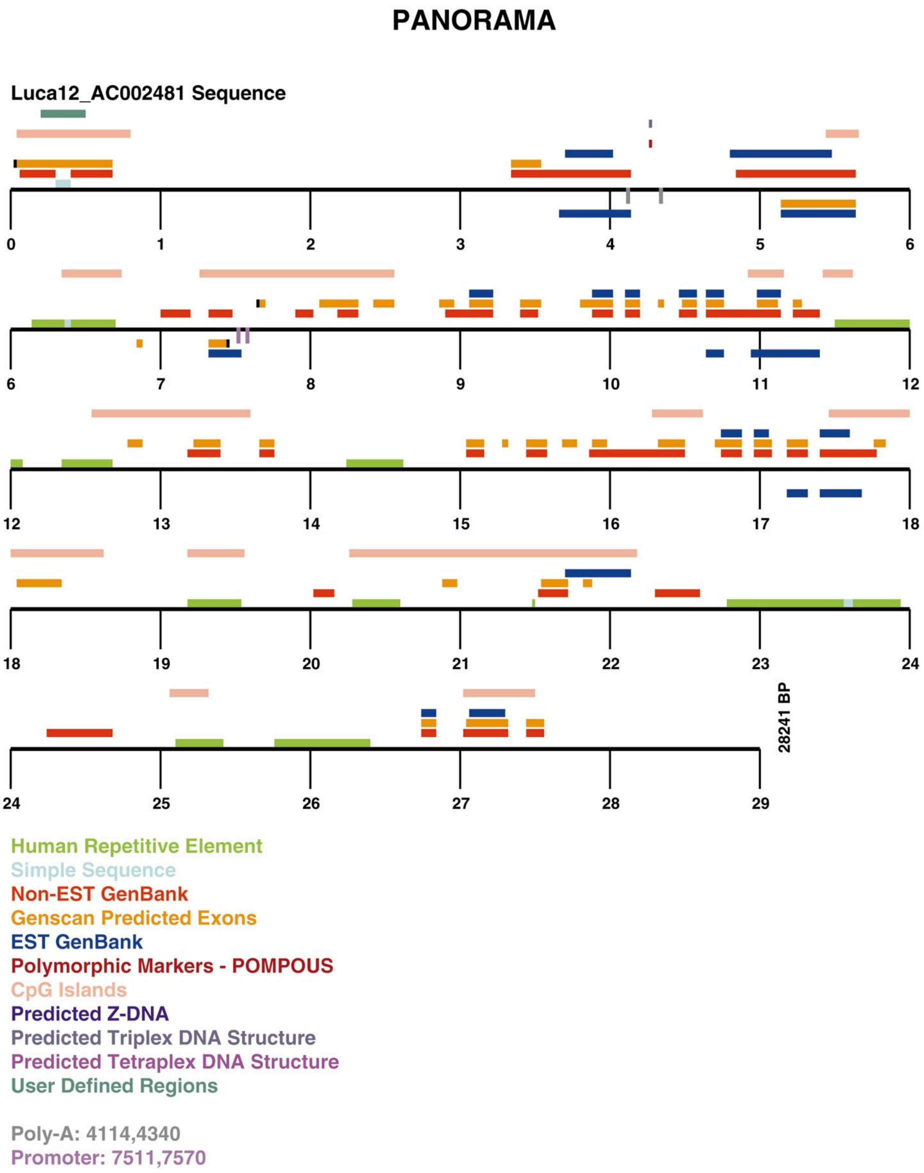


FIG. 4. PANORAMA plot for the input of 28,241 bp of LUCA 12 (AC002481). In the second line, the rightmost eight hits against the non-EST GenBank database correspond to parts of gene 21 (AF040707).

RESULTS

In the current intermediate gene discovery period (until completely annotated, finished sequence is available), most laboratories are involved in an iterative process involving informatics and "wet lab" work. One example of how we used PANORAMA in this effort resulted in the discovery of the voltage-gated calcium channel regulatory subunit CACNA2D2 (Gao *et al.*, 2000). The utility of PANORAMA is demonstrated through the analysis of 650 kb of genomic sequence from the human chromosome 3p21.3 region. This region has been associated with allele loss in lung cancer, breast cancer, and other cancers and is currently being studied for the identification of a tumor suppressor gene(s) that might be mutated in these cancers (Wei *et al.*, 1996).

Results from an analysis of cosmid LUCA 11 (Z84492) indicated hits in the EST GenBank database, but no open reading frames. The EST clones were used as probes for a Northern blot that yielded a 5.7-kb mRNA. The ESTs were then used to screen a cDNA library that returned a 3-kb partial cDNA clone, which when sequenced was found to have homology to the partial open reading frame of a voltage-gated calcium channel regulatory subunit. This appeared to be the 3′ end of the gene.

PANORAMA results for cosmids LUCA 6 (Z84493) and LUCA 7 (Z84494, ~100 kb centromeric of the LUCA 11 sequence) showed a few exon predictions by GenScan but no hits within the EST GenBank. The portions of the sequence representing the predicted exons were obtained and used as probes for reverse transcriptase-PCR. The RT-PCR yielded a cDNA product, which when sequenced yielded a cDNA corresponding to the exons predicted. The cDNA was then used as a probe for a Northern blot that returned a 5.7-kb mRNA, similar to that obtained from the LUCA 11 (Z84492)-related sequence. At the time of this analysis, the sequences of cosmids LUCA 6 (Z84493) and LUCA 11 (Z84492) were not completed and could not be assembled into one sequence. Once the cDNAs from the putative 3′ and 5′ ends of the new gene were sequenced, they were used as probes to rescreen the cDNA library for clones with both 3′ and 5′ ends. This allowed the construction of the full-length cDNA. PANORAMA was run on this cDNA, and the results were compared to those from LUCA 6 (Z84493), LUCA 7 (Z84494), LUCA 8 (Z84495), LUCA 9 (Z75743), LUCA 10 (Z75742), and LUCA 11 (Z84492). This helped determine the intron–exon boundaries and confirm the 5′-3′ orientation of the gene. The intron–exon boundary sequences were also used for genomic screening and mutation analysis by single-strand conformation polymorphism. The CpG islands predicted were used to identify the promoter regions. The combined results identified CACNA2D2 (gene 26, Accession No. AF040709), a candidate human tumor suppressor gene homologous to a voltage-gated calcium channel $\alpha_2\delta$ subunit. The gene is 5482 bp in length with 37 exons and spans 150 kb across cosmids LUCA 6, 7, 8, 9, 10, and 11 (Gao *et al.*, 2000). CACNA2D2 shows 58% homology and 57% similarity to the voltage-gated calcium channel $\alpha_2\delta$ subunit in *Mus musculus* and *Rattus norvegicus*, respectively. A similar analysis led to the identification of gene 21 (Accession No. AF040707), also a candidate human tumor suppressor gene, contained in cosmid LUCA 12 (Accession No. AC002481; Figs. 2, 3, and 4). It has 20 exons spanning 10 kb. It is 39% homologous to the *Saccharomyces cerevisiae* nitrogen permease regulatory subunit. It also shows 29% similarity to the *Caenorhabditis elegans* nitrogen permease regulator 2.

DISCUSSION

PANORAMA should be of use in the processes of gene discovery and sequence analysis. It precisely locates each of the characteristics of the sequence, provides a visual representation of the results, and allows the user to identify relationships between these features at a glance. PANORAMA also provides detailed results for each of the features identified in the form of its associated tables. Since PANORAMA is implemented as a modular set of programs and scripts, it is being continuously refined and expanded.

The utility of PANORAMA is demonstrated by its role in the discovery of two new genes, gene 21 (Accession No. AF040707) and CACNA2D2 (Accession No. AF040709) in addition to the analysis of already known genes in the 3p21.3 region of the human chromosome.

An important advantage of PANORAMA over other applications is that it is available on the World Wide Web (at http://atlas.swmed.edu). The results are displayed as a page with links to the various files generated, including the text results of the analysis, the PDF and PostScript static images, and the interactive JAVA applet. The availability of PANORAMA over the Internet provides the researcher access to a powerful sequence analysis and annotation tool.

ACKNOWLEDGMENTS

We thank Hewlett-Packard for providing servers on which a large portion of this work was performed. This work was supported by the Whitaker Foundation, the P. O'B. Montgomery Distinguished Chair, National Cancer Institute Grants CA71618 and Lung Cancer SPORE P50 CA70907, the G. Harold and Leila Y. Mathers Charitable Foundation, and NIH Grant P50 HG01002-01. M.I.L. and M.H.W. (NCI) were funded in whole with Federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. NO1-CO-56000. The content of the publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. M.H.W. is now at Celera Genomics, Rockville, Maryland.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389–3402.

Andrade, M. A., Brown, N. P., Leroy, C., Hoersch, S., de Daruvar, A., Reich, C., Franchini, A., Tamames, J., Valencia, A., Ouzounis, C., and Sander, C. (1999). Automated genome sequence analysis and annotation. *Bioinformatics* **15:** 391–412.

Bailey, L. C., Jr., Fischer, S., Schug, J., Crabtree, J., Gibson, M., and Overton, G. C. (1998). GAIA: Framework annotation of genomic sequence. *Genome Res.* **8:** 234–250.

Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268:** 78–94.

Cherepanov, P., Este, J. A., Rando, R. F., Ojwang, J. O., Reekmans, G., Steinfeld, R., David, G., De Clercq, E., and Debyser, Z. (1997). Mode of interaction of G-quartets with the integrase of human immunodeficiency virus type 1. *Mol. Pharmacol.* **52:** 771–780.

Fondon, J. W., III, Mele, G. M., Brezinschek, R. I., Cummings, D., Pande, A., Wren, J., O'Brien, K. M., Kupfer, K. C., Wei, M. H., Lerman, M., Minna, J. D., and Garner, H. R. (1998). Computerized polymorphic marker identification: Experimental validation and a predicted human polymorphism catalog. *Proc. Natl. Acad. Sci. USA* **95:** 7514–7519.

Gao, B., Sekido, Y., Maximov, A., Saad, M., Forgacs, E., Latif, F., Wei, M. H., Lerman, M., Lee, J. H., Perez-Reyes, E., Bezprozvanny, I., and Minna, J. D. (2000). Functional properties of a new voltage-dependent calcium channel $\alpha_2\delta$ auxiliary subunit gene (CACNA2D2). *J. Biol. Chem.* **275:** 12237–12242.

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. *J. Mol. Biol.* **196:** 261–282.

Harris, N. L. (1997). Genotator: A workbench for sequence annotation. *Genome Res.* **7:** 754–762.

Harris, N. L. (2000). Annotating sequence data using Genotator. *Methods Mol. Biol.* **132:** 259–276.

Jurka, J. (1998). Repeats in genomic DNA: Mining and meaning. *Curr. Opin. Struct. Biol.* **8:** 333–337.

Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992). CpG islands as gene markers in the human genome. *Genomics* **13:** 1095–1107.

Li, P., Davies, C. J., North, D., Schilling, P., Evans, G. A., and Garner, H. R. (1997a). Supercomputing in genomic sequencing: Optimization of BLAST and other sequence algorithms for high speed parallel processing. *Sci. Comput. Automation* **14:** 19–24.

Li, P., Kupfer, K. C., Davies, C. J., Burbee, D., Evans, G. A., and Garner, H. R. (1997b). PRIMO: A primer design program that applies base quality statistics for automated large-scale DNA sequencing. *Genomics* **40:** 476–485.

Murakami, K., and Takagi, T. (1998). Gene recognition by combination of several gene-finding programs. *Bioinformatics* **14:** 665–675.

Scharf, M., Schneider, R., Casari, G., Bork, P., Valencia, A., Ouzounis, C., and Sander, C. (1994). GeneQuiz: A workbench for sequence analysis. *Proceedings of the 2nd International Conference on Intelligent Systems for Molecular Biology* **2:** 348–353.

Sinden, R. R. (1994). “DNA Structure and Function,” Academic Press, San Diego.

Smith, R. F., Wiese, B. A., Wojzynski, M. K., Davison, D. B., and Worley, K. C. (1996). BCM Search Launcher—An integrated interface to molecular biology data base search and analysis services available on the World Wide Web. *Genome Res.* **6:** 454–462.

Staden, R. (1996). The Staden sequence analysis package. *Mol. Biotechnol.* **5:** 233–241.

Thomas, M. J., Freeland, T. M., and Strobl, J. S. (1990). Z-DNA formation in the rat growth hormone gene promoter region. *Mol. Cell. Biol.* **10:** 5378–5387.

Ulrich, M. J., Gray, W. J., and Ley, T. J. (1992). An intramolecular DNA triplex is disrupted by point mutations associated with hereditary persistence of fetal hemoglobin. *J. Biol. Chem.* **267:** 18649–18658.

Wei, M. H., Latif, F., Bader, S., Kashuba, V., Chen, J. Y., Duh, F. M., Sekido, Y., Lee, C. C., Geil, L., Kuzmin, I., Zabarovsky, E., Klein, G., Zbar, B., Minna, J. D., and Lerman, M. I. (1996). Construction of a 600-kilobase cosmid clone contig and generation of a transcriptional map surrounding the lung cancer tumor suppressor gene (TSG) locus on human chromosome 3p21.3: Progress toward the isolation of a lung cancer TSG. *Cancer Res.* **56:** 1487–1492.