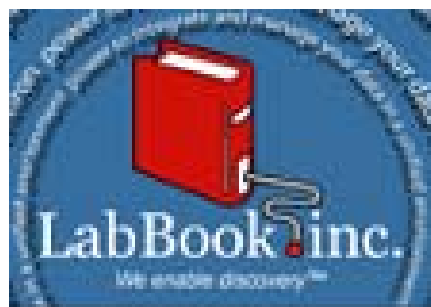


バイオ分野におけるXMLの応用事例

~BSMLとXML Genomic Suite DB™のご紹介~

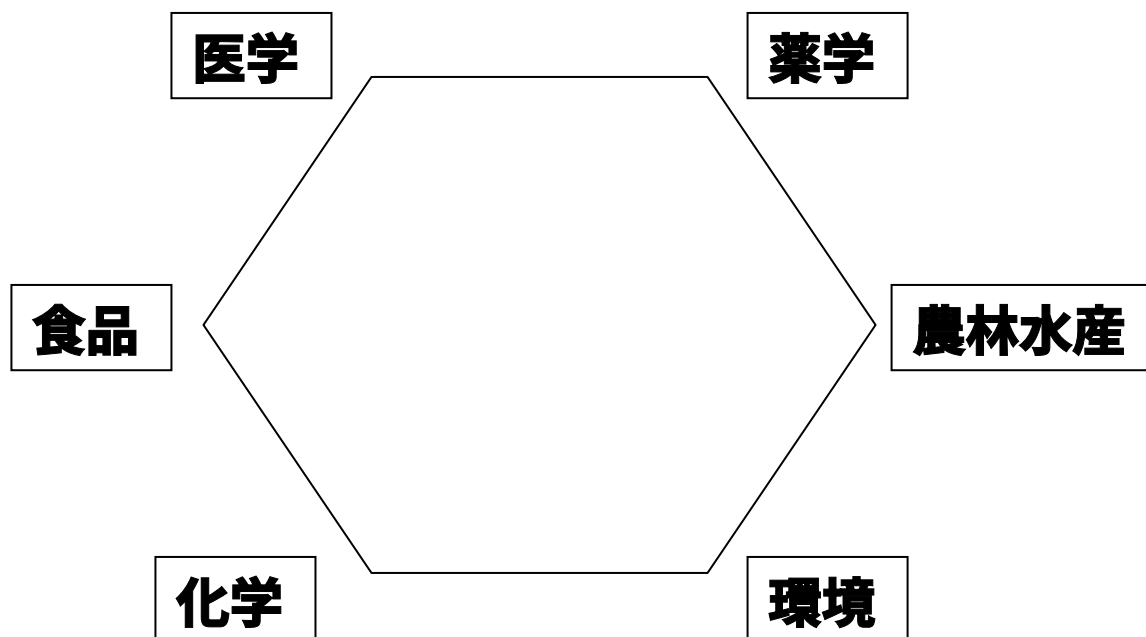


株式会社CRCソリューションズ
応用システム事業部
バイオインフォマティクス推進チーム
安中 敏男

発表内容

- バイオ分野におけるXMLの動向
- 遺伝子配列データベースについて
- 遺伝子配列データについて
- BSMLのご紹介とその特徴
- ゲノム創薬とバイオインフォマティクスについて
- LabBook社のご紹介
- BSMLを用いた製品のご紹介

1. バイオ応用分野の専門領域



各種のゲノムプロジェクト

各種のゲノムデータベース

2. バイオ分野におけるXMLの現状

1) 配列データ・アノテーション (データベース) 分野

BSML、GAME、BIOML、DDBJ-XML、TIGR XML、AGAVE、BSA、NCBI XMLなど

2) 遺伝子発現 (マイクロアレイ) 分野

GeneXML、GEML、MAML、MAGE-ML

3) 文献情報分野

MEDLINE

他にも、オントロジー、生物モデル (生物シミュレータ)、分子など様々な分野で、XML化が進められている。

3. 遺伝子配列データベースの特徴

- 1) 各データの大きさがまちまち (3桁以上の隔たり) である。
→ RDBでは、取扱いが困難。XMLではデータの大きさには無関係。
- 2) 大量データの更新が生じやすい。
→ XMLを用いるとデータの入れ替えが容易。
- 3) データエントリー数が非常に多く (DDBJでは14Mエントリー)、現在も増加しつつあり、検索時間がかかる。
→ XML化により、タグによる検索項目の絞込み、XML形式のフラットファイルを直接検索でき、高速化される。
- 4) データのやりとりがWebベースである。
→ データのフラットファイル化が望ましい。

⇒ 遺伝子配列データベースのXML化には大きなメリットがある。

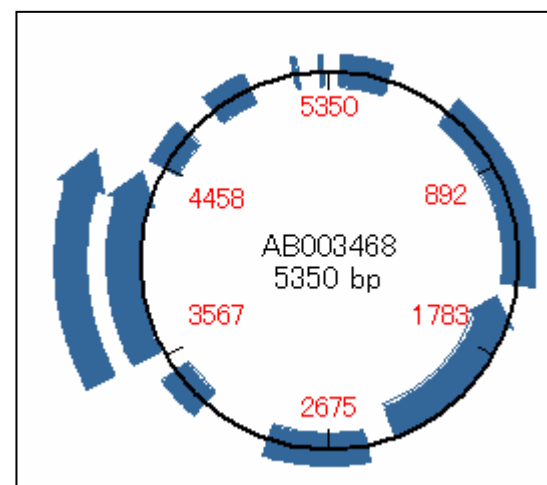
4. 配列データを取り扱うときの問題 (その1)

1) 配列データを検索しても、記号の羅列が表示されるだけである。

→ 理解しやすい表示方法が必要

```

ORIGIN
1  ggtaaccttct gaggcggaaa gaaccagccg gatccctcga gggatccaga catgataaga
61  tacatttgatg agtttggaca aaccacaaact agaattcagt gaaaaaaatg ctttattttgt
121 gaaaattttgt atgctattgc tttattttgta accattataa gctgcaataa acaagttaac
181 aacaacaatt gcatttcattt tatgtttcag gttoaggggg aggtgtggga ggttttttaa
241 agcaagtaaa accctacaaa atgttggtatg gotgattatg atccggctgc ctccggcggt
301 tccgtgatga cgggtgaaaa ctctgacaca tgcagctccc ggagacgggc acagcttgtc
361 tgtaagcgga tgcgggggag agacaagccc gtcaggggcg gtcagcggtt gttggcggtt
421 gtcggggcgc agccatgacc cagtcacgta gcgatagcgg agtgtatact ggcttaacta
481 tgcggcatca gacgagattg taactgagag goaccatatt cgggtgtgaa taaccgacag
541 atgogtaagg agaaaatacc gcatcaggcg ctcttcggct tctctgctca ctgactgctt
601 gcgctcggtc gttcggctgc ggcgagcggt atcagctcac tcaaaaggcg taatacgggt
661 atccacagaa tcagggggata acgcaggaaa gaacatgtga gcaaaaagcc agcaaaaaggc
721 caggaaacgt aaaaaggcgc cgttgcctgc gtttttccat aggcctccgc cccctgacga
781 gcatcacaaa aatcgacgct caagtcagag gtggcgaaac ccgacaggac tataaagata
841 ccaggcggtt ccccttgtaa gctccctcgt gcgctctcct gttccgaccc tgcggttac
901 cggataacct tcgcgcttgc tcccttcggg aagcgtggcg ctttctcata gctcagctgc
961 taggtatctc agttcgggtg aggtcgttgc ctccaagctg ggcgtgtgtc acgaaccccc
1021 cgttcagccc gaocgctgcg ccttatccgg taactatcgt cttgagtcca acccggttag
1081 acacgactta tcgccactgg cagcagccac tggtaacagg attagcagag cgaggatagt
1141 aggcgggtgt acagagttct tgaagtgtgt gcctaaacta ggcctacacta gaagaacagt
1201 atttggtatc tgcgctctgc tgaagccagt taacctcgga aaaagagttg gtagctcttg
1261 atccggcaaa caaaaccacc ctggtagcgg tgggtttttt gtttgcaagc agcagattac
1321 ggcgagaaaa aaaggatctc aagaagatcc tttgatcttt tctacggggt ctgacgctca
1381 gtggaacgaa aactcagctt aagggttttt ggtcatgaga ttatcaaaaa ggcatttcaa
1441 ctgatacttt ttaaatataa aatgaagttt taaatcaatc taaagtatat atgagtaaac
1501 ttggtctgac agttaccaat gcttaatcag tgaggcaact atctcagcga tctgtctatt
1561 tcttcatccc atagttgcct gactccccgt cgtgtagata actacgatac gggaggcgctt
1621 accatctggc cccagtgtgc caatgatacc gcgagaccca cgtcaccggc ctccagattt
  
```



4. 配列データを取り扱うときの問題 (その2)

2) 配列データを解析するには、記号の羅列は不適當である。

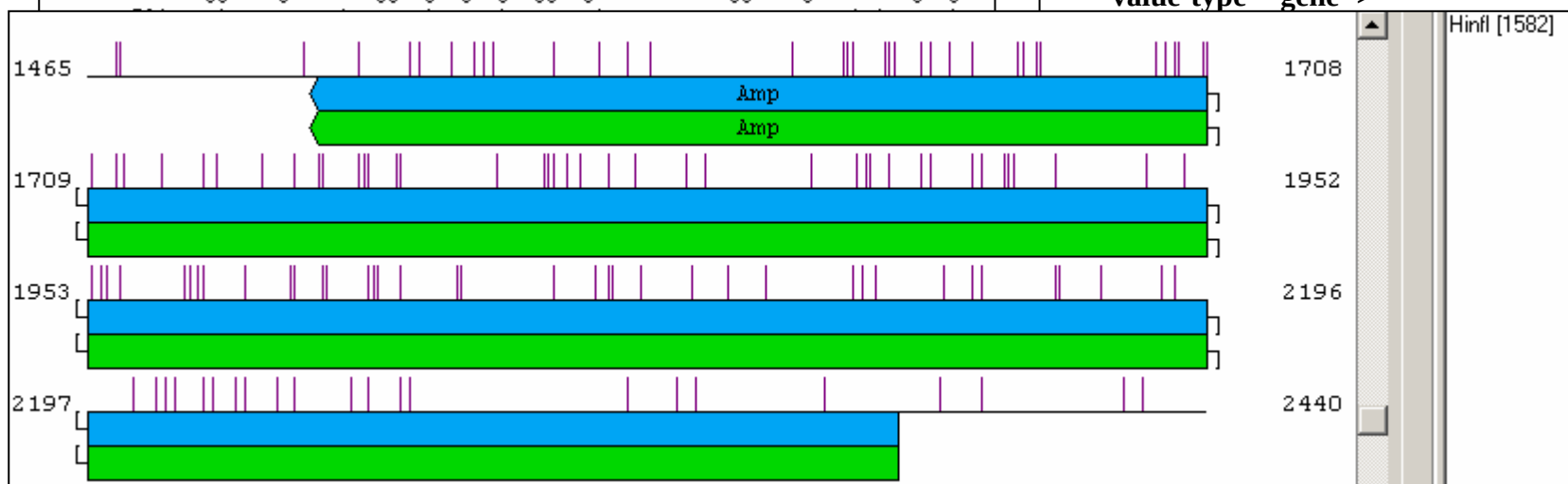
→ 解析しやすいデータ表現方法が必要

```

ORIGIN
1  ggtaaccttct gaggcggaaa gaaccagccg gatccctcga gggatccaga catgataaga
61  tacattgatg agtttggaca aaccacaact agaattgcagt gaaaaaaatg ctttatttgt
121 gaaatttgtg atgctattgc tttatttgtta accattataa gctgcaataa acaagttaac
181 aacaacaatt gcatttcattt tatgtttcag gttoaggggg aggtgtggga ggttttttaa
241 agcaagtaaa acctctacaa atgttggtatg gotgattatg atccgctgc ctcgcgcgtt
301 tcggtgatga cggtgaaaao ctctgacaca tgcagctccc ggagacggtc acagcttgtc
361 tgtaagcgga tgcggggagc agacaagccc gtcaggcgcc gtcagcgsgt gttggcgsgt
421 gtcggggcgc agccatgacc cagtcaagta gcgatagcgg agtgtatact ggcttaacta
481 tgcggcatca gaggcagattg taactgagagt gcaccatatt cgggtgtgaaa taaccgcacag
541 atgogtaagg agaaaatacc gcatcaggcg ctcttccgct tctctgcctc ctgactcgt
601 gcgctcggtc gttcggctgc ggcgagcggt atcagctcac tcaaaaggcg taatacgggt
661 atccacagaa tcaggggata acgcaggaaa gaacatgtga gcaaaaagcc agcaaaaaggc
721 cagggaacct aaaaaggccg cgttgcctgc gtttttccat aggcctccgc cccctgacga
  
```

<Restriction-enzyme offset="1" overhang="3"
sites="594,669,1065,1582,2656,2678,4210,4344,4
396,5195" title="HinfI"/>

<Feature class="GENE" display-auto="1"
id="AB003468.FT2.FR4" title="Amp"
value-type="gene">



5. (XML準拠の) BSMLの有効性

1) 階層構造表現

- ・ 研究データそのものの階層構造化
- ・ 配列データ (例: フィーチャ、クオリファイアなど) の階層表現化

2) タグによる配列情報の意味付け

- ・ 膨大な検索量の低減化、高速化、検索ノイズの低減化
- ・ XMLブラウザ (Genomic XML Browser及びGenomic XML viewer) による理解しやすい表示

3) Open Standard

- ・ 他のDTDの追加により汎用化・最適化が可能
- ・ ASN.1 (Abstract Scientific Notation) とコンパチブル
- ・ フォーマットコンバータによるインテグレーションが可能

6. ゲノム創薬の流れ

- 1) 疾患関連遺伝子の解明
- 2) 疾患関連タンパクの解明
- 3) タンパクの立体構造の解明
- 4) 化合物のデザイン
- 5) 新薬の化学合成・治験・臨床

特徴

多種・多様・多量な実験データ
大規模な計算機シミュレーション
実験データやプロトコルなどの厳重な管理

バイオインフォマティクス



- ① 多種・多様なフォーマットで、かつ多量なデータ管理
- ② インターネットでの検索と検索結果の管理
- ③ ゲノム情報の理解しやすい表示機能
- ④ データの一括管理・共有化

7. LabBook社とBSML(Bioinformatic Sequence Markup Language) について

1) LabBook社



- ・ 1999年設立、従業員数約50名
- ・ 本社米国オハイオ州コロンバス、支社カリフォルニア州バークレーなど
- ・ I3C発足当時からのメンバー
- ・ BMSなどの世界的な製薬企業やNASA、オハイオ州立大学などの研究機関と共同研究も推進している。またWeilyなどの情報サービス企業とも提携している。

2) BSML



- ・ LabBook社が提案しているバイオインフォマティクス (特に配列とアノテーション) に特化したXML準拠のフォーマット。
- ・ 従来は理解しにくかった配列情報が非常に理解しやすく表示できるようになった。
- ・ EMBL (European Molecular Biology Laboratory) など有名なデータベースがBSMLをサポートしており、BSMLがゲノム配列データベース分野の標準になりつつある。

8. LabBook社が提供するソリューション

① eLabBook™

XML技術を用い、多様なフォーマットのデータを一括管理するツール
GLP、GCP基準に基づく実験ノート機能
インターネットにシームレスにリンクし、簡単に様々な検索が可能
研究者ごとの共有化できるデータの設定が可能
独自のデータベースの作成と統合も可能

② Genomic XML Browser™

BSMLにより配列情報を理解しやすく表示するビューワ
複雑なゲノムデータベースの検索が非常に簡単
新たなアノテーションの付加も可能

③ OHGDB-Query™

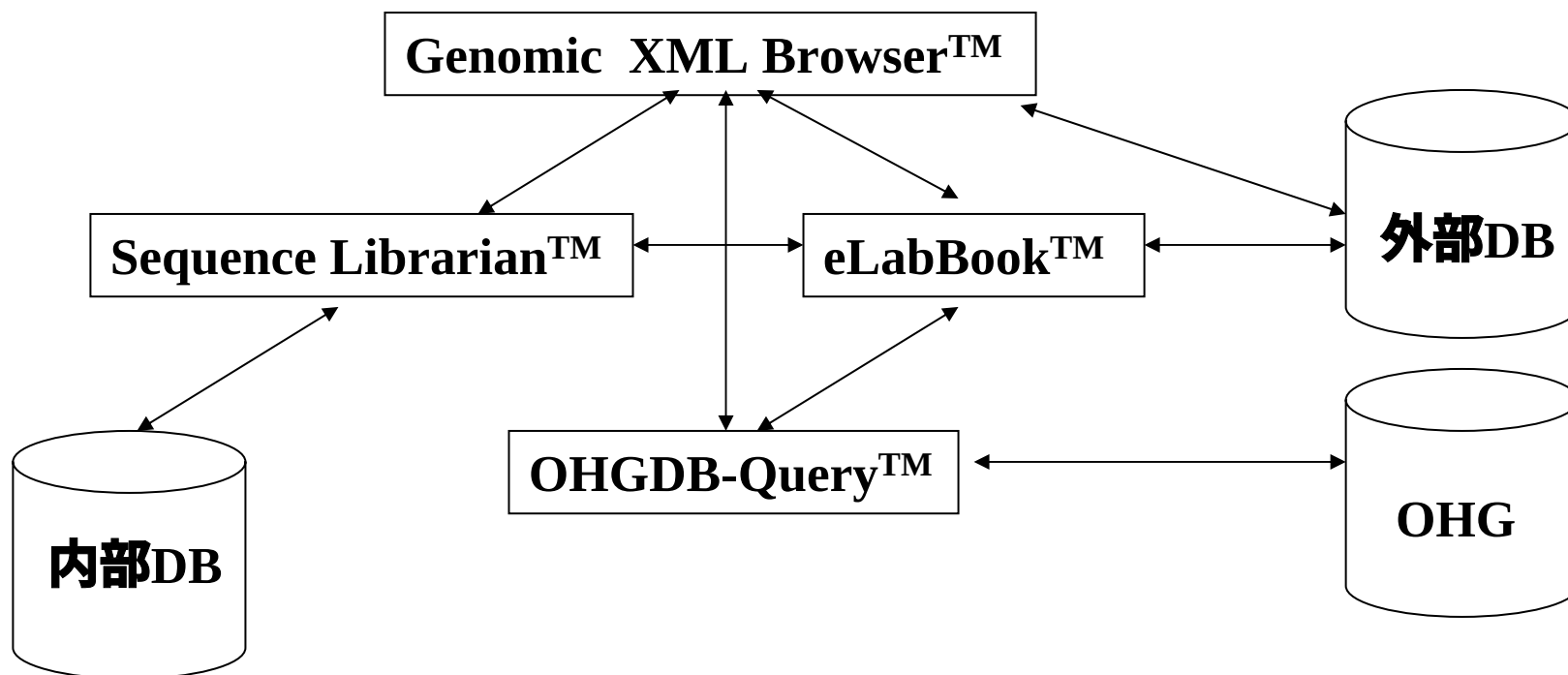
オハイオ州立大学のスパコンを用いて作られたヒトゲノムデータベース
にアクセスするためのインターフェース

④ Sequence Librarian™

独自のゲノムデータベース作成のためのビルダ

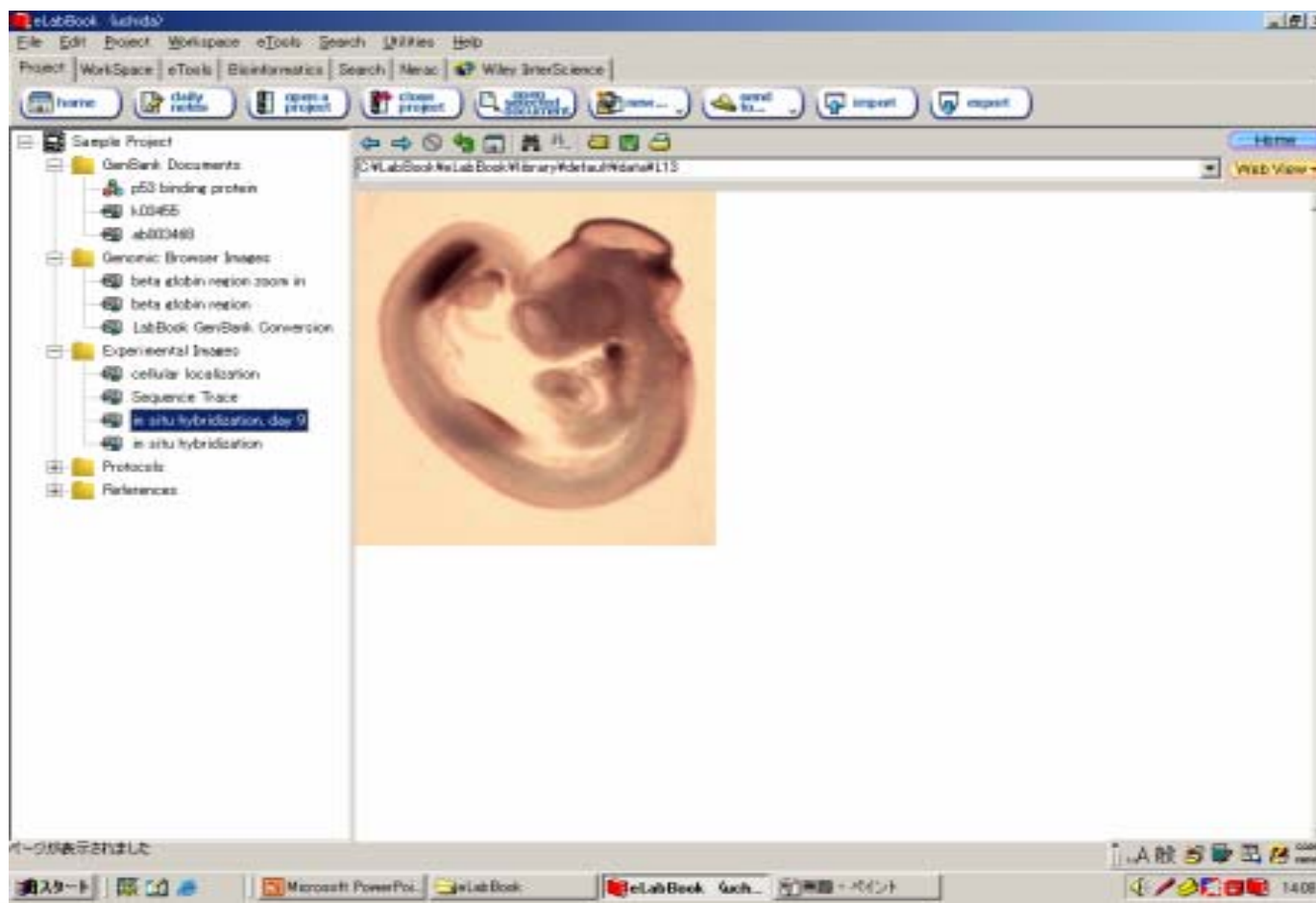
XML Genomic Suite DB™

9. XML Genomic Suite DB™の構成



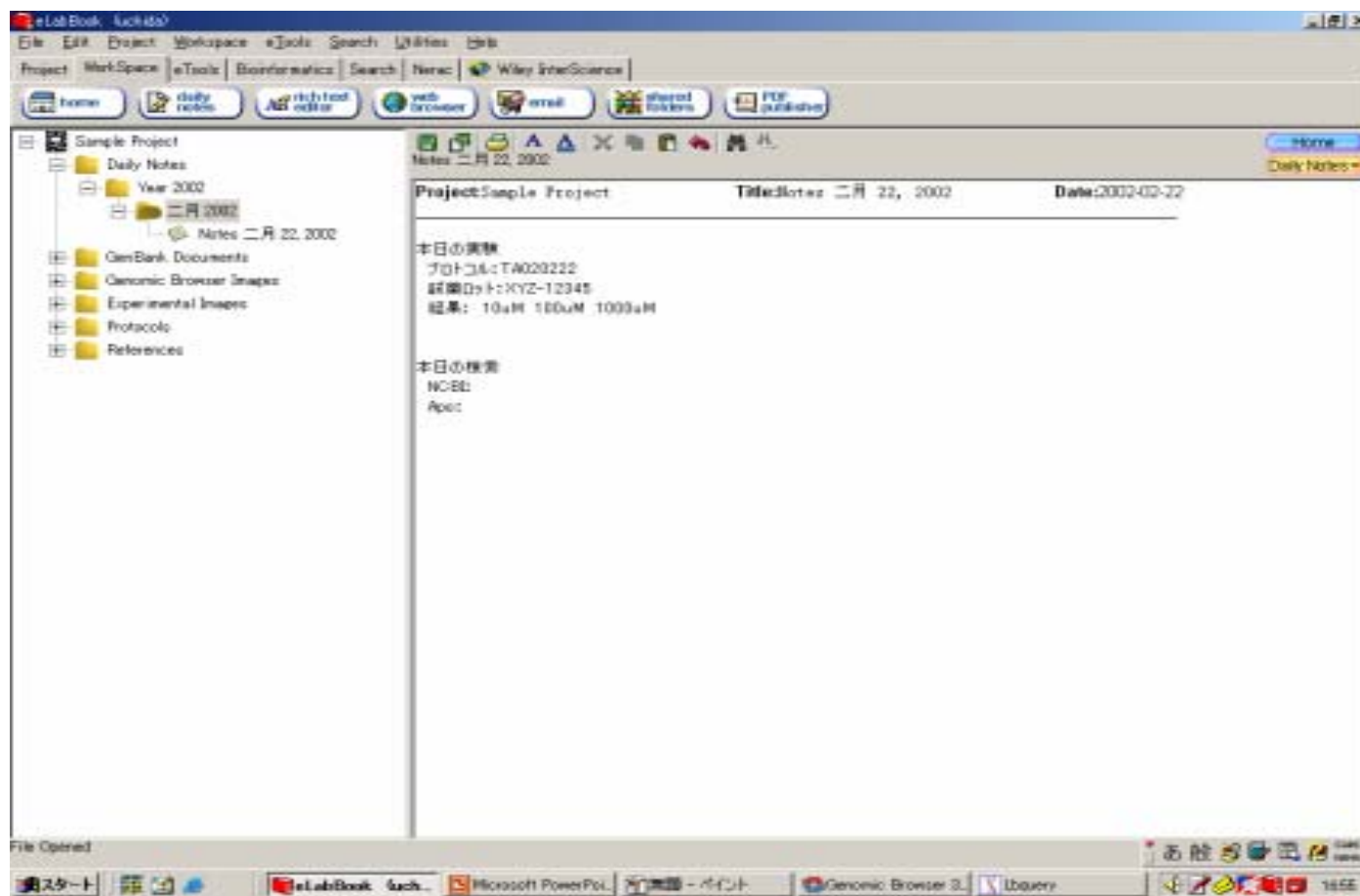
10. eLabBook™のご紹介 (1)

XML技術を用い、多様なフォーマットのデータを一括管理するツール



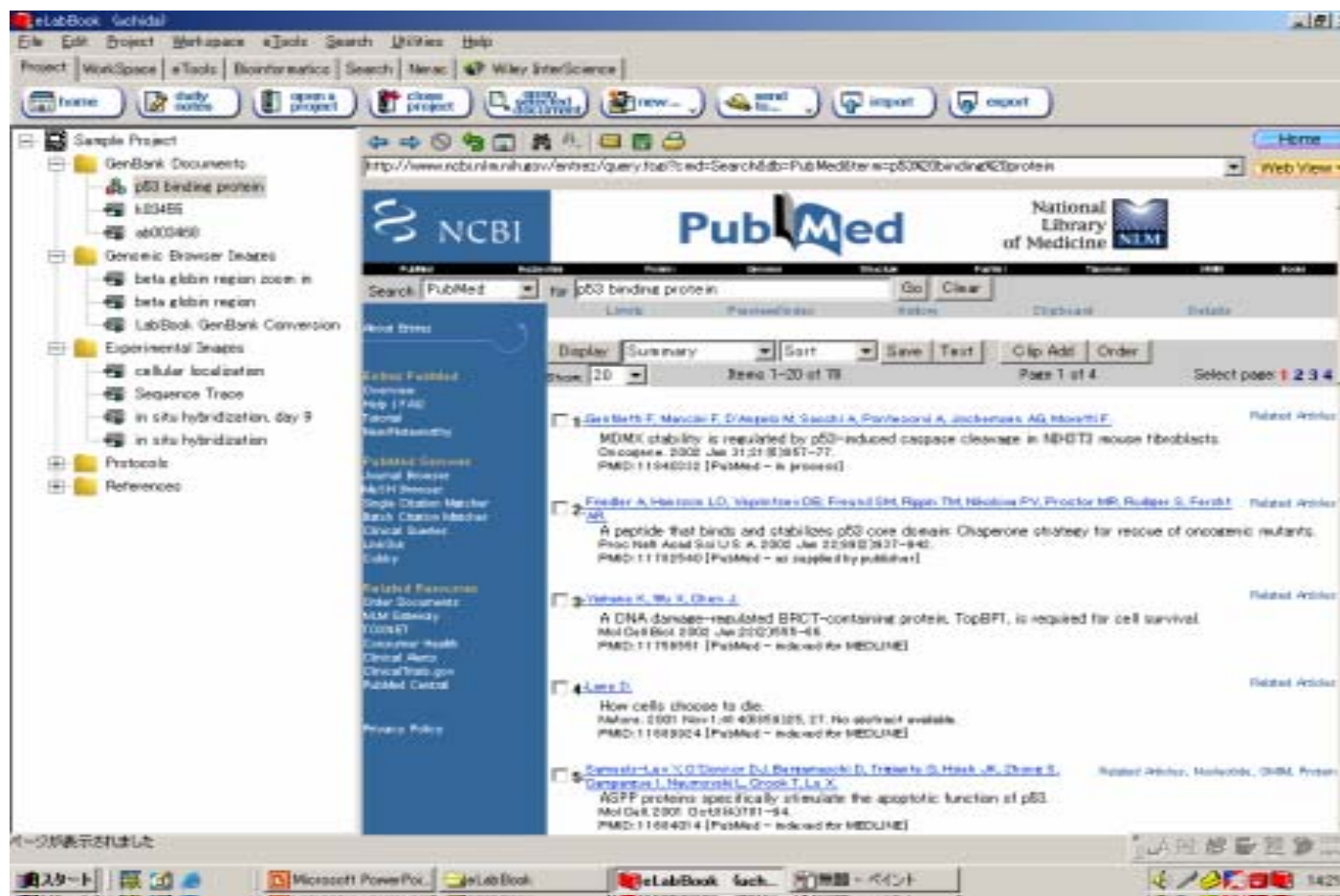
10. eLabBook™のご紹介 (2)

GLP、GCP基準に基づく実験ノート機能



10. eLabBook™のご紹介 (3)

インターネットにシームレスにリンクし、簡単に様々な検索が可能



The screenshot displays the eLabBook application window. On the left is a 'Sample Project' tree with folders like 'GenBank Documents', 'Generic Browser Images', 'Experimental Snapshots', 'Protocols', and 'References'. The main area shows a web browser window with the NCBI PubMed search results for 'p53 binding protein'. The search results list several articles, including 'MDM1 stability is regulated by p53-induced caspase cleavage in MPT3 mouse fibroblasts' and 'A peptide that binds and stabilizes p53 core domain: Chaperone strategy for rescue of oncogenic mutants'.

10. eLabBook™のご紹介 (4)

研究者ごとの共有化できるデータの設定が可能



Folder Contents: norip

Folder Owner: Noritaka Uchida

--select folder action--

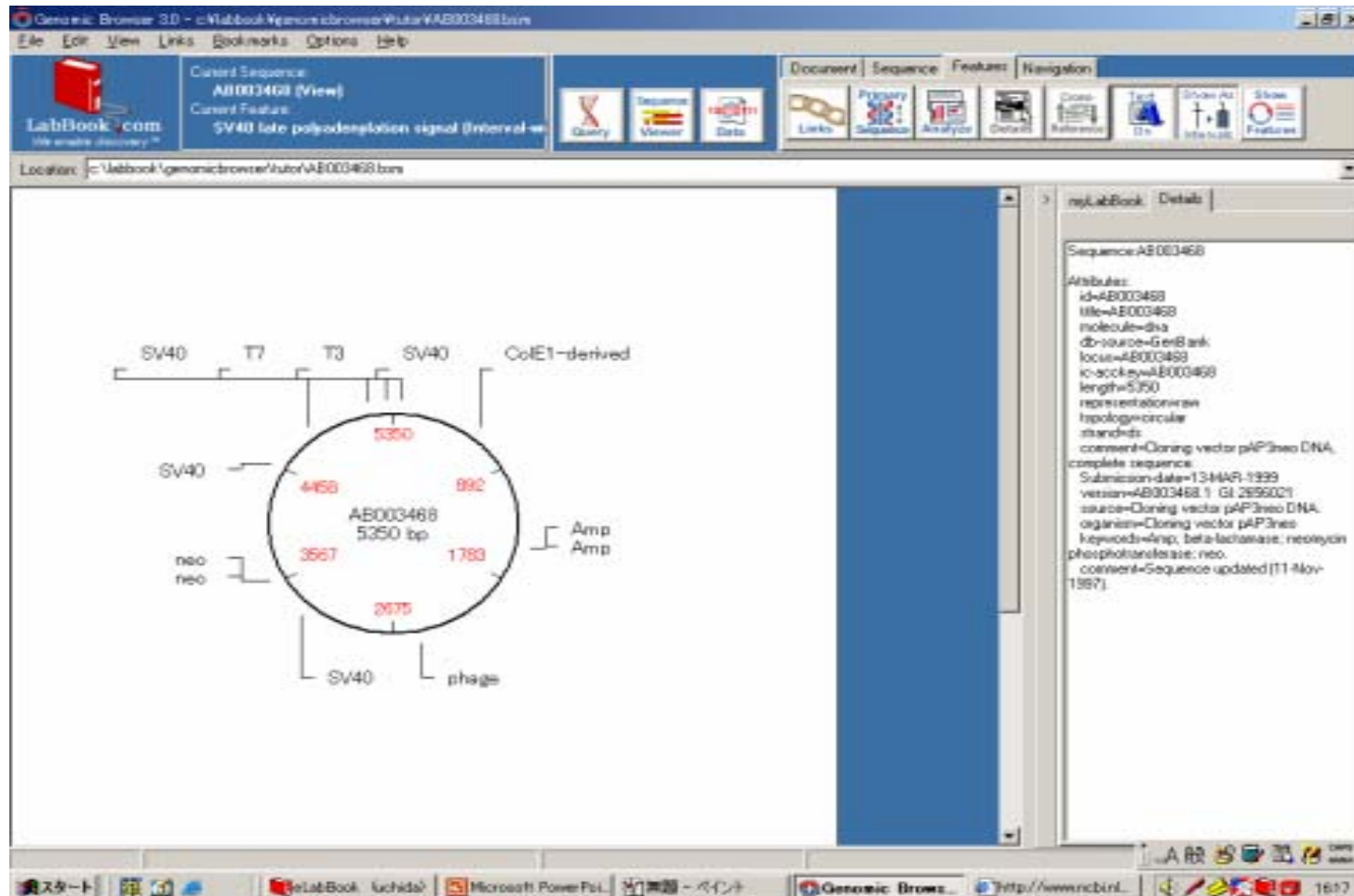


Document Name	Action	Type	Version	Size (kb)	Comments	Last Updated	Updated By	In PDF Queue?
3Dstructure.gif	--document action--	(gif)	1	22	new file upload	1/11/2002 4:37:31 AM	Yusuke Tanabe	no
LabBook NIKKEI English.doc	--document action--	(doc)	1	21	new file upload	1/29/2002 11:13:29 PM	Noritaka Uchida	no



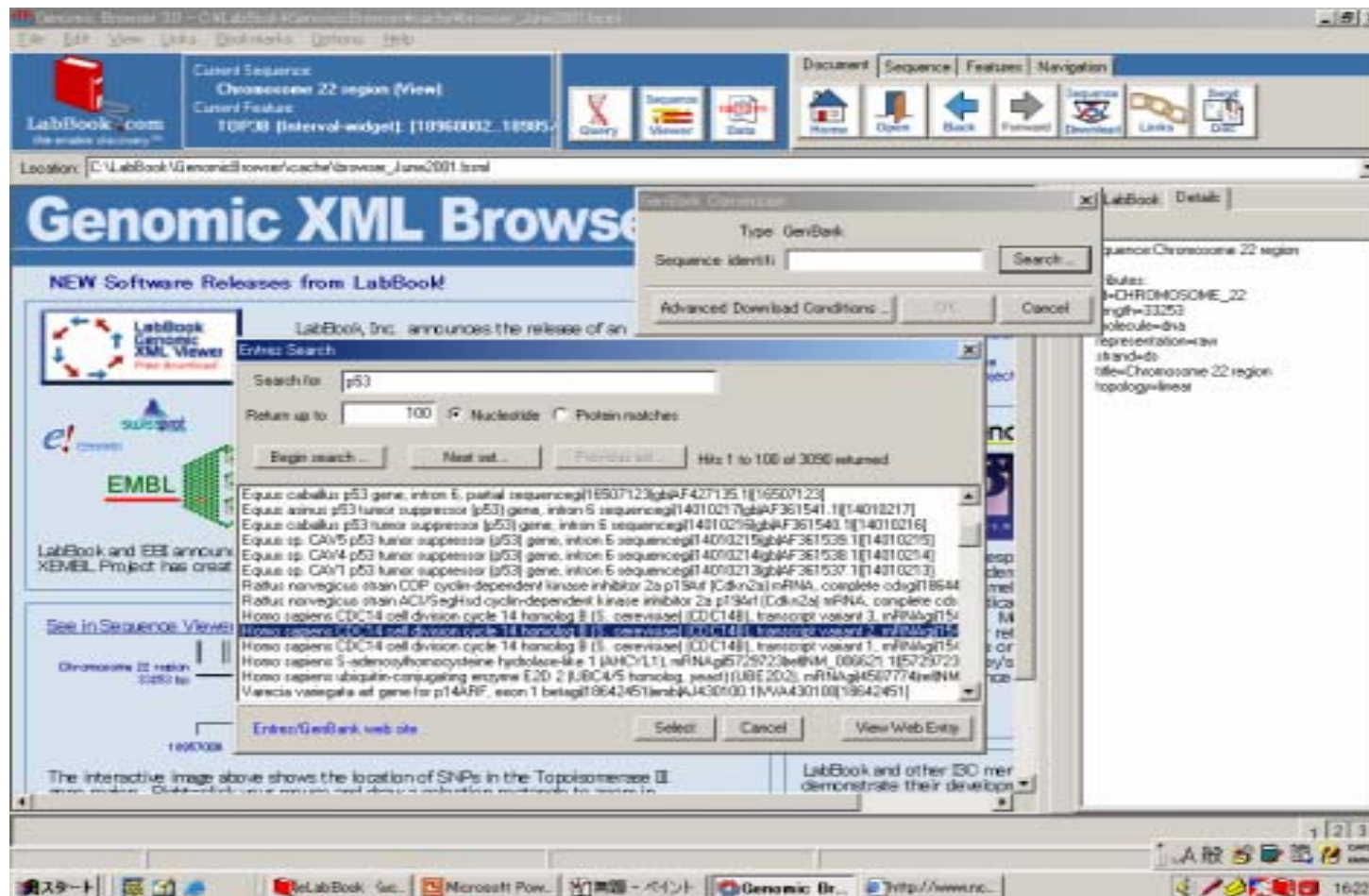
11. Genomic XML Browser™のご紹介 (1)

BSMLにより配列情報を理解しやすく表示するビューワ



11. Genomic XML Browser™のご紹介 (2)

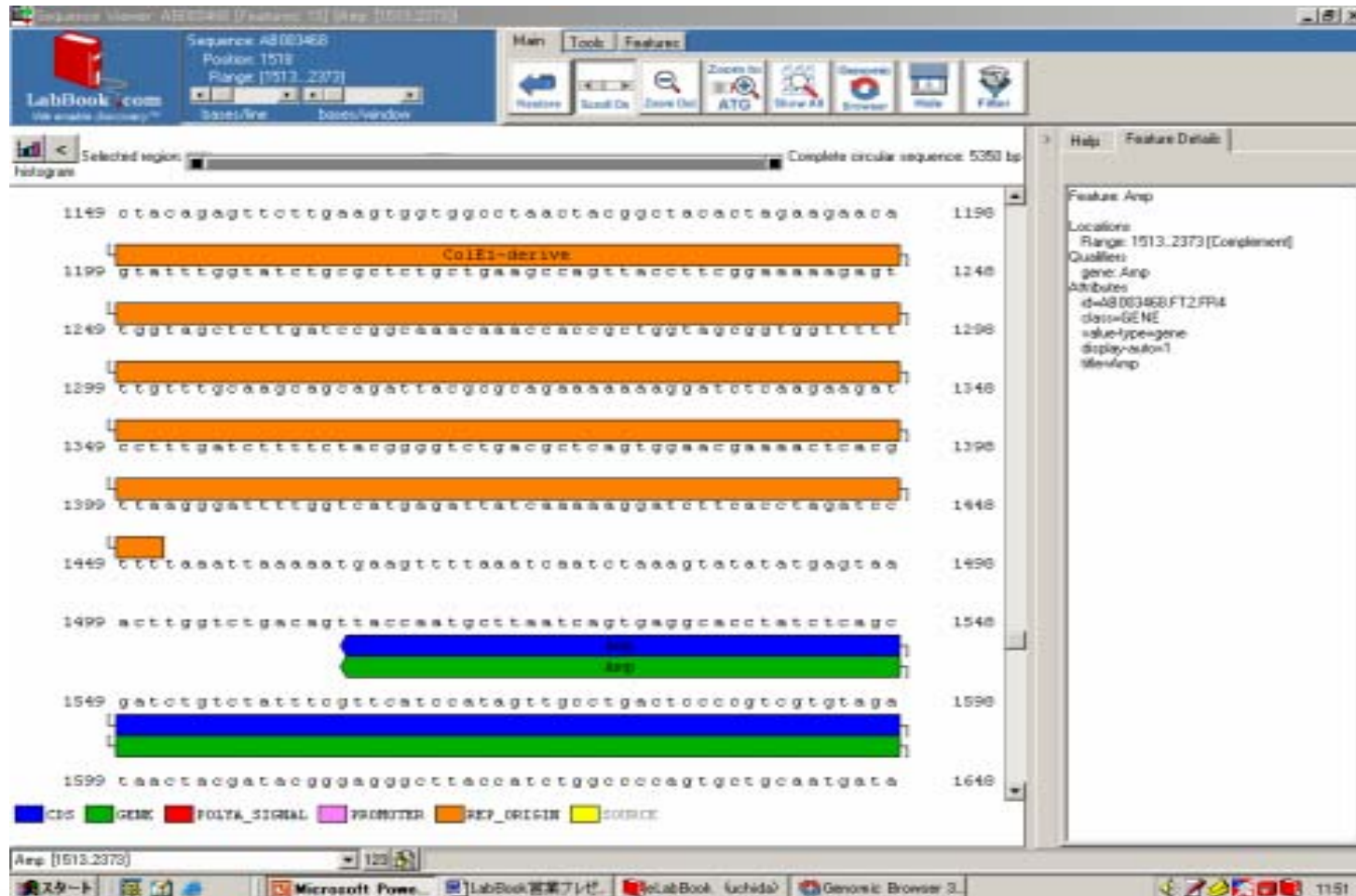
複雑なゲノムデータベースの検索が非常に簡単



The screenshot displays the Genomic XML Browser interface. At the top, there's a navigation bar with tabs for Document, Sequence, Features, and Navigation. Below this, a search bar is visible with the text "p53" entered. A list of search results is shown, including entries for "Equus caballus p53 gene, intron 5, partial sequence" and "Homo sapiens p53 tumor suppressor (p53) gene, intron 5 sequence". The interface also features a sidebar with links to various databases like EMBL and a main content area displaying genomic data for Chromosome 22.

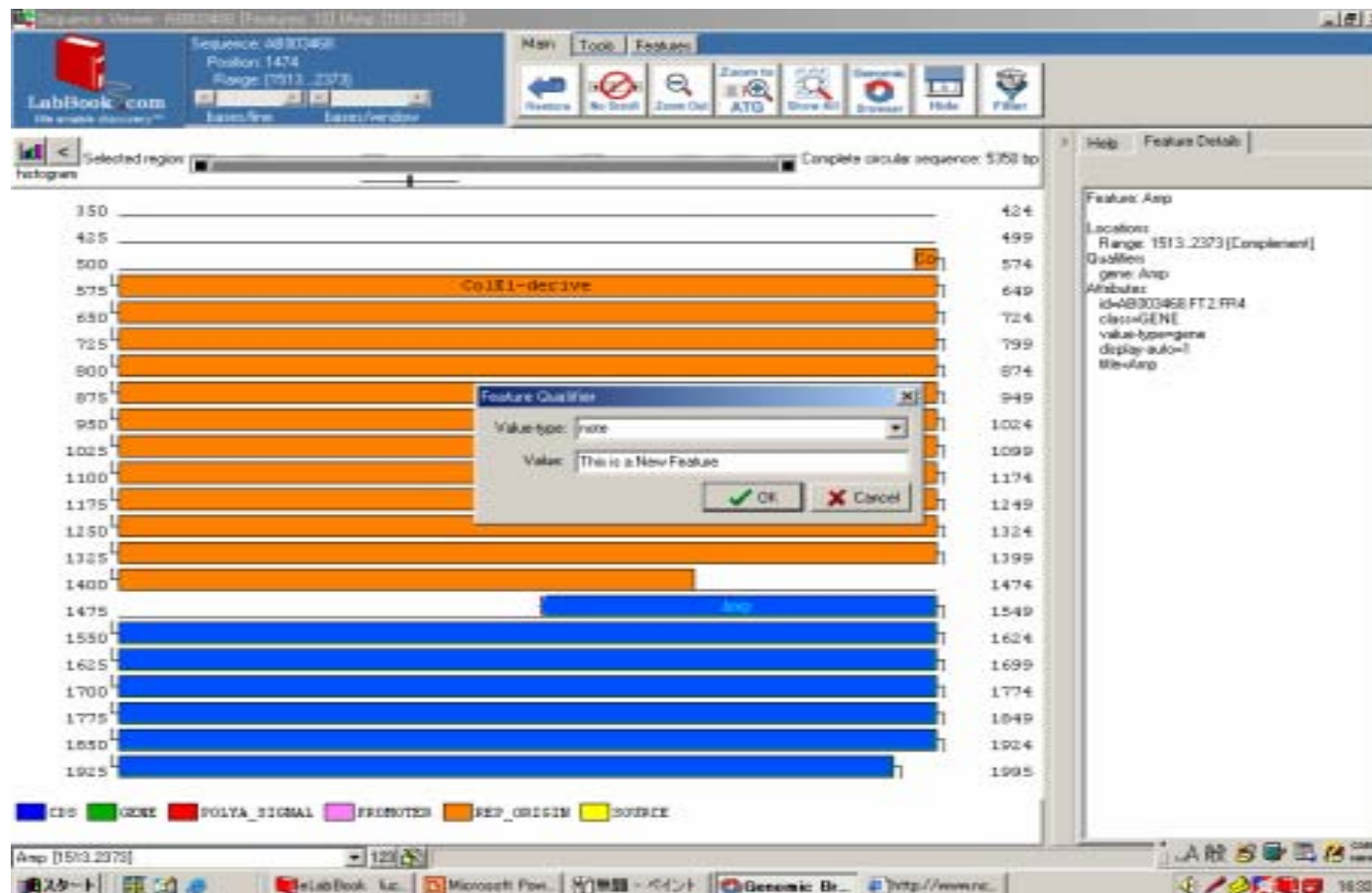
11. Genomic XML Browser™のご紹介 (3)

様々な細かさで配列データを表示することが可能



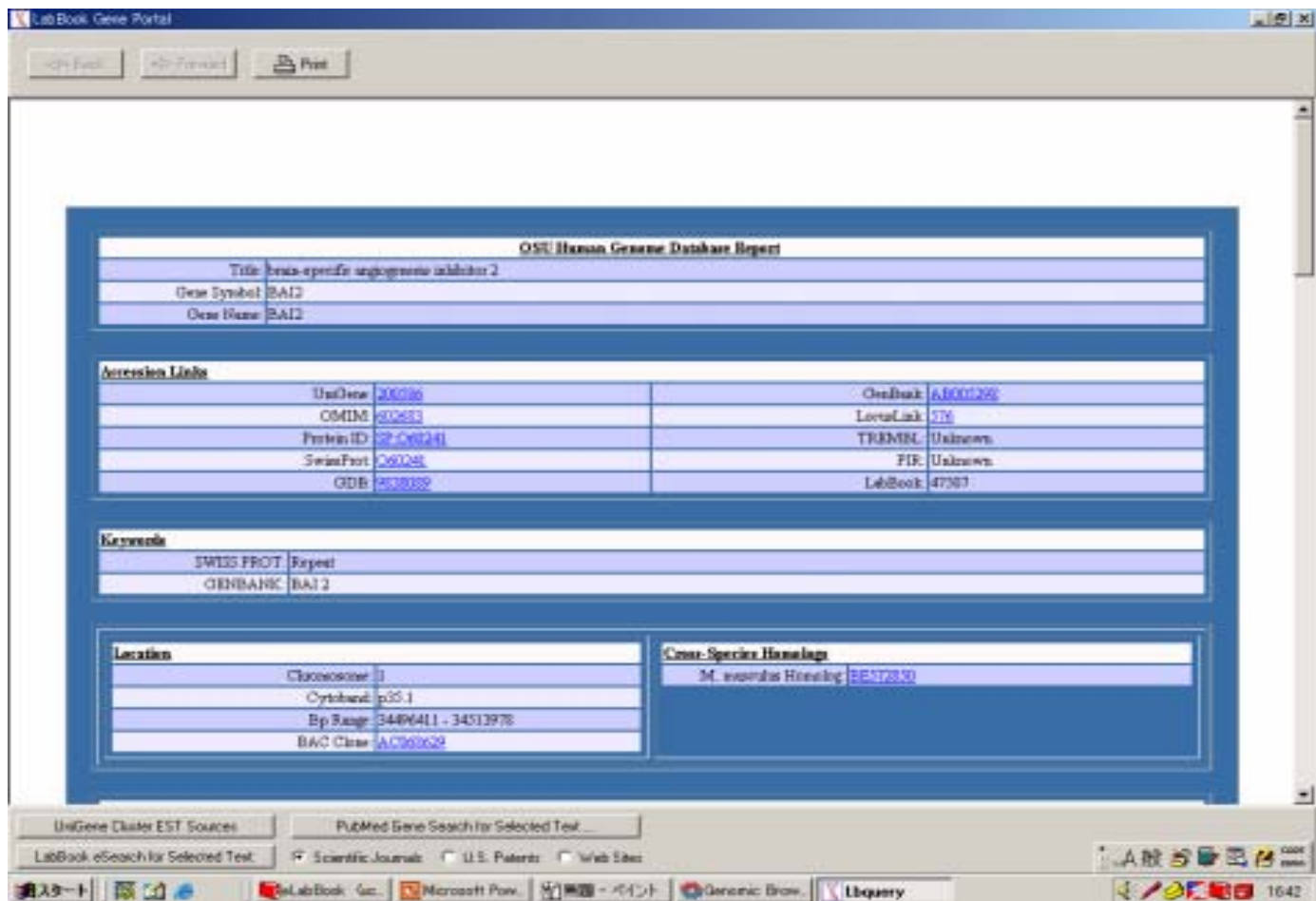
11. Genomic XML Browser™のご紹介 (4)

新たなアノテーションの追加や修正が可能



12. OHGDB-Query™のご紹介

オハイオ州立大学のヒトゲノムデータベースへのインターフェース



The screenshot displays the 'LabBook: Gene Portal' window with a report titled 'OSU Human Genome Database Report'. The report details the gene BAI2, including its title, symbols, and various accession numbers. It also provides a location map and cross-species homology information.

OSU Human Genome Database Report	
Title:	brain-specific angiogenesis inhibitor 2
Gene Symbol:	BAI2
Gene Name:	BAI2

Accession Links	
UniGene	200786
OMIM	602683
Protein ID	P2-C00241
SwissProt	C00241
QDB	Q03030
GeneBank	AB000295
LocalLink	JY
TRIMIN	Unknown
FIR	Unknown
LabBook	47307

Keywords	
SWISS PROT	Excerpt
GENBANK	BAI2

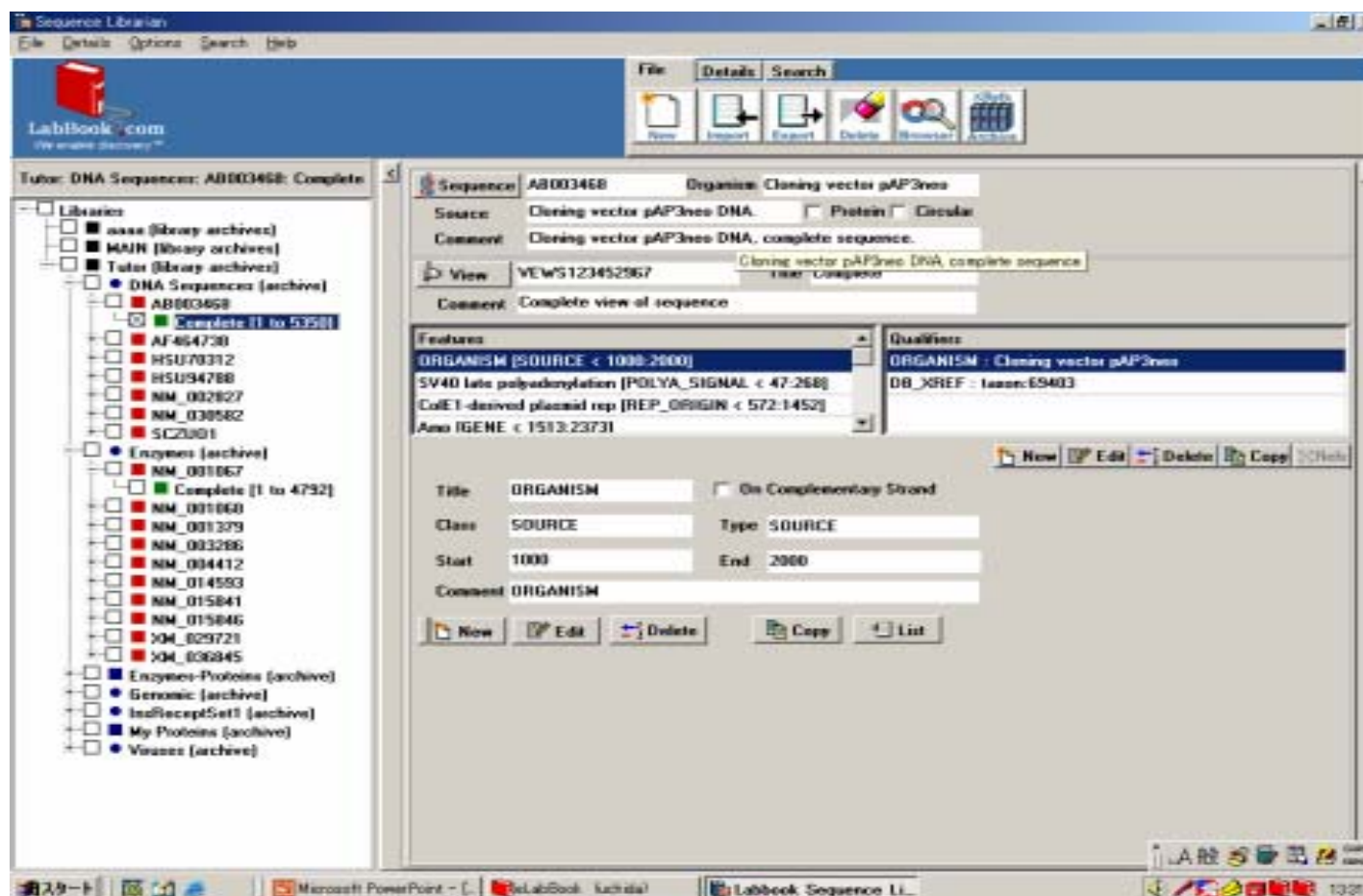
Location	
Chromosome	1
Cytoband	p35.1
Bp Range	3489411 - 34513978
BAC Clone	AC090029

Cross-Species Homology	
M. musculus Homology	EE572830

At the bottom of the window, there are search buttons: 'UniGene Cluster EST Sources', 'PubMed Gene Search for Selected Test...', 'LabBook eSearch for Selected Test', and 'Scientific Journals'. The taskbar shows the system clock at 16:42 and the date 10/2/2001.

13. Sequence Librarian™のご紹介

ユーザ独自の配列データベースビルダ



14. 将来の製品ラインアップ

- 1) プラットフォームに依存しないJavaバージョン
- 2) 企業内の閉じた環境で実行できるエンタープライズバージョン



BSML 2.2DTDなどの各種のリソースは

**<http://www.bsml.org/Resources/default.asp>
からダウンロードできます。**

また、ASN.1⇔BSMLのコンバータソースコードと変換例は

**<http://www.bsml.org/>
からダウンロードできます。**



Genomic XML Browser™の機能を絞った体験版 (Genomic XML Viewer™)は
http://www.labbook.com/products/viewer_register.asp
から無料で入手できます。是非、お試しください。



また、本製品のお問い合わせは
販売元: CTCラボラトリーシステムズ株式会社
電話 03-3419-9171 e-mail: biosupport@ctc-g.co.jp



または
技術サポート: 株式会社CRCソリューションズ
電話 03-5634-5829 e-mail: bio@crc.co.jp



までお願い申し上げます。