

# 유전자 진화 분석을 위한 전유전체 가시화 Component Ware 개발

조치영\*, 박수현<sup>1</sup>, 김대수<sup>2</sup>, 하홍석<sup>2</sup>, 안궁<sup>2</sup>, 김희수<sup>2</sup>, 조환규<sup>1</sup>  
 부산대학교 {U-Port 정보기술사업단\*, 컴퓨터공학과<sup>1</sup>, 생물학과<sup>2</sup>}  
 cycho@pusan.ac.kr, shpark@pearl.cs.pusan.ac.kr,  
 {kds2465, hay2k, ankung79, khs307, hgcho}@pusan.ac.kr

## Development of Whole Genome Visualization Component Ware for Evolutionary Genetic Analysis

Chi-Young Cho\*, Soo-Hyun Park<sup>1</sup>, Dae-Soo Kim<sup>2</sup>, Hong-Seok Ha<sup>2</sup>, Kung Ahn<sup>2</sup>,  
 Heui-Soo Kim<sup>2</sup>, Hwan-Gue Cho<sup>1</sup>  
 {Center for U-Port IT Research and Education\*, Dept. of Computer Science and Engineering<sup>1</sup>,  
 Division of Biological Sciences<sup>2</sup>} Pusan National University

### 요 약

많은 양의 유전자 정보가 유전공학의 발전과 Genome 프로젝트의 결과로 축적되고 있으며, 이러한 유전자 정보를 체계적으로 관리하고 가시화하기 위한 생물정보학 분야의 연구가 진행되고 있다. 기존의 많은 Genome Browser들이 완성된 형태의 툴로 서비스되고 있다. 이러한 툴들은 다목적의 많은 기능을 포함하고 있어 특정연구를 진행해야하는 연구자들은 너무 많은 정보로부터 원하는 것을 찾기 위해 시간과 노력이 필요하게 된다. 본 논문에서는 특정한 목적의 Gene 가시화 툴을 제작할 수 있는 Component Ware를 제안하고 이를 이용한 진화분석용 가시화 툴을 소개한다.

### 1. 서 론

유전공학의 발전과 최근까지의 연구결과로 방대한 크기의 유전정보가 축적되고 있다. 이러한 유전자 정보를 효율적으로 탐색하기 위해 생물정보학 분야의 연구가 진행되고 있고 유전자 정보를 탐색하고 가시화하는 다목적의 Gene Browser들이 개발되고 있다[1,2,3,4,5]. 표 1에 Gene Browser의 예를 보이고 있다.

표 1: Genome Browser 예

Browser	특징	참고
VISTA	GenBank data의 human, mouse alignment 결과 및 유전 정보 제공	[1]
ECR	사용이 간편한 웹 기반 genome browser를 통한 척추동물 genome alignment 및 분석 정보 제공	[2]
PipMaker	대용량 DNA 서열 비교 및 가시화	[3]
Combo	FASTA, GenBank 등 annotation 가시화 및 Annotation에 따른 Whole genome 비교	[4]
UCSC	대표적인 genome browser로서 총 32종의 다양한 유전자 정보 분석 제공	[5]

이러한 Gene Browser들은 많은 기능을 포함하고 있어 개별 연구를 진행해야 하는 연구자들은 너무 많은 정보로 인해 오히려 자신이 원하는 것을 찾는데 어려움을 겪

을 수 있다. 본 논문에서는 특정 유전자들이 종의 진화 과정에서 나타나는 변화를 분석할 수 있는 Component Ware를 제안하고 이를 이용해 생성한 유전자 진화분석용 Gene 가시화 툴을 소개한다. 유전자 가시화는 그 유전자가 존재하는 구간 주변에서 엑손(exon)과 인트론(intron)의 위치를 블록 사각형과 라인의 형태로 선형적으로 표시되어질 수 있다. 그림 1은 본 논문에서 제안한 Component Ware를 사용해 CDC23과 KIF20A라는 유전자를 가시화한 것이다.

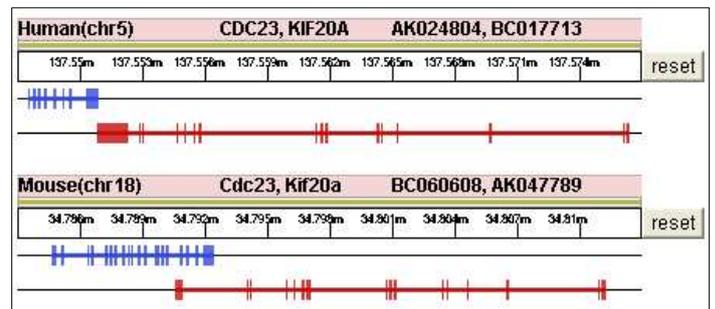


그림 1: Gene Display 예

이 두 유전자는 Human의 5번 염색체와 Mouse의 18번 염색체에서 그림과 같이 일부 부분이 Overlap되어 나타나고 있으며 Human과 Mouse에 모두 존재한다. 이와 같이 어떤 유전자들이 서로 다른 종에 길이와 위치가 다소 차이를 보이며 존재할 수 있는데 본 논문에서는 이와

같이 특정 유전인자들이 종(species)간 어떤 변화를 가지는지 분석할 수 있는 유전자 가시화 Component Ware를 소개하고 이를 이용해 생성한 유전자 가시화 툴을 소개한다.

UCSC, Vista 등 많은 Genome Browser들이 웹 브라우저를 사용한 접속을 지원하고 있다. 이는 기존의 대부분의 운영체제에서 웹 브라우저를 지원하고 있으므로 이를 이용한 쉬운 접속을 지원하기 위한 것이다. 본 논문에서 제안하는 Component Ware에서도 웹 브라우저를 사용한 접속을 지원하여 이 Component Ware로 생성된 프로그램이 클라이언트의 운영체제나 웹 브라우저에 독립적으로 구동될 수 있도록 한다. 또한 Gene 정보가 방대하므로 Database 사용은 필수적이다. 본 논문에서 제안하는 Component Ware에서 DBMS로 MySQL을 사용하고 있으나 다양한 DBMS를 사용할 수 있도록 설계되었다. 클라이언트는 DBMS와의 접속을 직접적으로 하지 않고 서버의 DBMS 연동 Agent와 TCP/IP Streaming을 함으로써 한 번 제작된 툴이 DBMS의 변경 때문에 재구성되어야 하는 것을 막을 수 있도록 했다. 구현 언어는 java언어를 사용한다. 시스템 실행에 필요한 JRE 자동배포는 MS Windows의 Internet Explorer (JRE 자동배포기능) 등 다양한 클라이언트 환경에서 수행될 수 있도록 하였다.

이 후 장에서의 내용은 다음과 같다. 제2장에서 Component Ware 클래스의 시스템 구성을 소개하고 3장에서는 본 논문에서 제안한 Component Ware의 주요 클래스의 특성을 소개한다. 제4장에서는 Component Ware 클래스를 사용해 개발된 Gene 가시화 도구의 클라이언트 특성을 소개하고 제5장에 결론을 제시하였다.

## 2. Component Ware 시스템 구성

### 2-1. Component Ware 개요

그림 2는 본 논문에서 제안한 Component Ware의 주요 클래스의 전체 구조와 데이터/제어 흐름을 나타내고 있다. 이 Component Ware의 클래스를 이용해 개별 Gene 가시화 도구를 개발할 수 있게 된다.

Component Ware의 주요 클래스는 크게 클라이언트 측과 서버 측 두 부분으로 구성된다. 클라이언트 측은 사용자 인터페이스 부분과 서버 시스템으로 접속을 실행하기 위한 부분으로 구성된다. 사용자 인터페이스를 담당하는 클래스들은 Gene 정보 표시, Gene 형태 출력, Gene 위치를 탐색하기 위한 Ruler 컨트롤 클래스 등 사용자에게 검색결과를 보여주거나 검색을 하기 위해 필요한 제어부분으로 이루어져 있다.

서버 시스템으로 접속을 하기 위한 부분은 Query Manager 클래스가 담당하며 사용자 인터페이스에서 사용될 쿼리 결과 저장용 클래스인 Query Result 클래스와 연동한다.

EvOG 주요 클래스 데이터/제어 흐름도

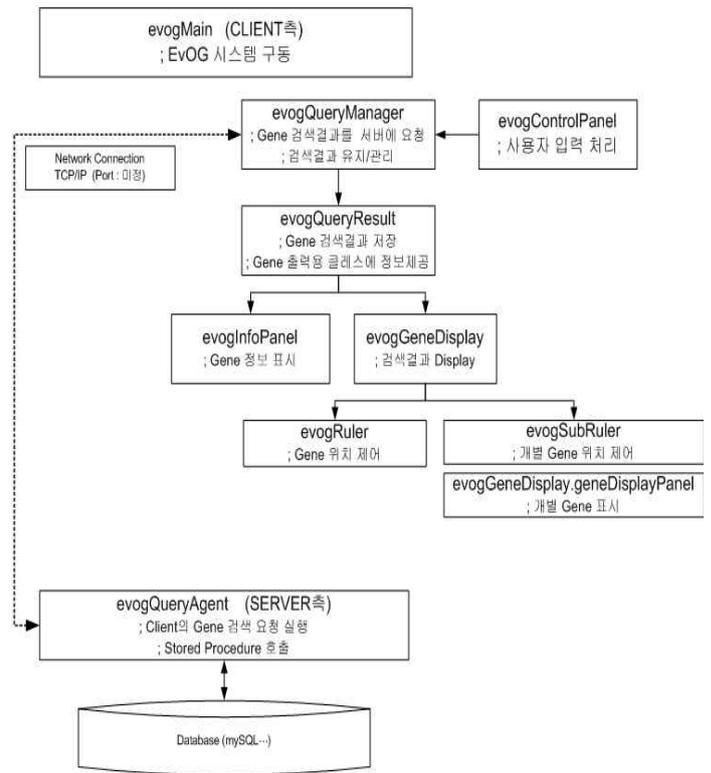


그림 2: Component Ware 데이터/제어 흐름도

그림 3은 클라이언트의 검색명령을 받아 DBMS와 연동을 수행하는 서버측을 나타내고 있다. 클라이언트의 Query Manager에서 내린 검색명령을 서버의 Query Agent가 받아들이고 서버에서 실행하고 클라이언트로는 결과만을 전송하도록 수행된다. 즉, DBMS와의 연동은 서버의 Query Agent만 담당하며 클라이언트의 Query Manager 클래스는 사용자가 내린 검색 명령만을 서버측의 Query Agent로 전달한다.

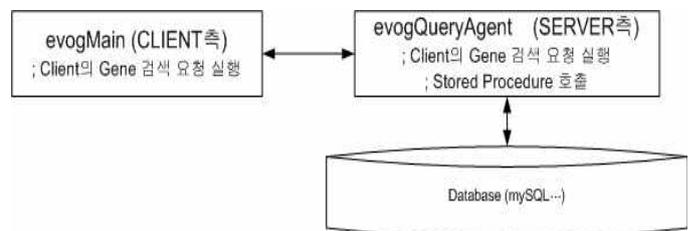


그림 3: Component Ware Core DBMS연동

Query Agent는 DBMS와 연동하여 검색한 결과를 문자열의 형태로 받아 Query Manager에 전달하는 역할을 수행한다. 클라이언트의 Query Manager는 서버의 DBMS의 종류가 어떤 것이든 상관없이 서버측의 Query Agent와 연동을 수행하도록 함으로써 DBMS 독립적으로 수행될 수 있도록 하였다. 이러한 3-Tier Client/Server

특성은 DBMS 독립적 Component Ware를 위한 필수적인 요소이다. 이러한 특성은 서버 측의 성능 및 기능 향상을 위한 코드 수정시에도 Client측은 독립적으로 수행될 수 있게 된다. 또한 Client가 Java Applet 환경이므로 클라이언트 측 프로그램의 업그레이드시에도 웹 브라우저로 접속시 자동적으로 다운로드되어 실행되므로 클라이언트측은 이러한 세부적 내용을 알 필요가 없게 된다.

### 3. Component Ware 주요 클래스

#### 3.1 Query Manager 클래스

사용자가 내린 검색명령에 따라 서버의 Query Agent 클래스와 연동하고 검색결과를 다운로드 받아 Query Result에 저장하는 역할을 수행한다. 이 클래스에서 수행하는 주요기능은 그림 4와 같다. 서버상태 확인과 서버와 TCP/IP 스트리밍 접속은 이 클래스의 생성자에서 수행된다. 이 클래스를 사용하는 곳에서 Query Manager 클래스의 객체 생성시 접속을 수행할 서버의 IP와 포트 번호, 사용자ID와 비밀번호를 지정하게 된다.

서버 상태 확인  
 서버와 TCP/IP 접속  
 유전자 검색 Query를 Query Agent에 전달  
 Query 결과를 Query Result 클래스에 저장

그림 4: Query Manager 클래스의 주요 기능

검색 Query는 개별 유전자를 검색하는 기능과 진화 분석을 위해 두 유전자로 검색하는 기능으로 나뉘며, 보다 세부적인 검색이 필요한 경우 사용자정의 Query를 실행할 수 있게 하였다. Query Manager의 주요 메서드의 프로토타입을 그림 5에 나타내었다. 이 메서드들의 실행결과는 Query Result에 저장된다.

```
public void QueryByGeneName(String geneName);
public void QueryByGeneName(String geneName1,
                             String geneName2);
public void QueryByPosition(String species,
                             String chr,
                             int start, int end);
public void QueryByUserDef(String queryString);
```

그림 5: Query Manager 클래스의 주요 메서드

#### 3.2 Query Result 클래스

Query Manager에 실행에 의해 서버의 Query Agent에서 보내오는 결과는 Query Result에 저장된다. 이 클래스에서는 검색 결과의 유전자 정보(유전자이름, 관련

종(species), 염색체, RefSeq정보, Exon구간 정보 등)와 검색 결과(총 레코드 수, 오류발생 여부 등) 및 실행 모드(단일 유전자 검색, 진화분석을 위한 유전자 검색)를 저장하고 Gene Display 클래스에서 이 정보들을 출력하게 된다. 그림 6은 Query Result 클래스의 주요 메서드와 멤버 변수들을 보이고 있다.

```
// gene 정보
Vector m_geneName, m_species, m_direct,
      m_mRNA, m_refSeq,
      m_chromosome;
int m_exe_mode;

// gene 위치정보
public Vector m_genomeStart, m_genomeEnd;
public Vector m_exon_start[];
public Vector m_exon_end[];

public int getRecordCount();
```

그림 6: Query Result 클래스의 주요 멤버와 메서드

#### 3.3 Gene Display 클래스

Gene Display 클래스는 Query Result 클래스에 저장된 유전자 정보들을 출력하는 기능을 수행한다. 개별 유전자 출력 기능과 진화 분석을 위해 두 유전자 출력 기능으로 구성된다. 그림 7과 그림 8은 검색된 유전자들을 출력하는 그림이다. 검색된 결과 중 상단부의 위치이동 Ruler 부위를 마우스 드레깅하여 부분 확대를 수행할 수 있다. 그림 7은 개별 유전자를 표시하는 기능이며, 특정 유전자가 존재하는 종(species)과 염색체번호를 같이 출력해주는 것을 보이고 있다.

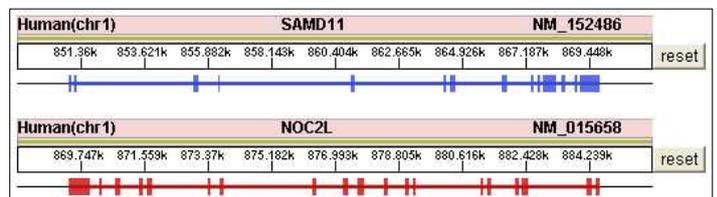


그림 7: 개별 유전자 출력

본 논문에서 사용한 유전자 정보는 Chicken, Chimp, Cow, Drosophila, Gambiae, Human, Mouse, Rat, Xtropicallis, Zebrafish의 10종으로 UCSC Gene Data [6]를 사용하였다. 그림 8은 어떤 두 유전자가 같은 염색체에 존재하는 종이 있는지 검색하여 그 두 유전자의 형태

를 출력해주는 그림이다. 그림에서 KIF20A와 CDC23 유전자가 Chimp, Human, Mouse, Rat에 존재하며 고등생물인 Chimp와 Human은 Overlap되는 곳이 비교적 작으며 Mouse와 Rat은 큰 것을 알 수 있다.

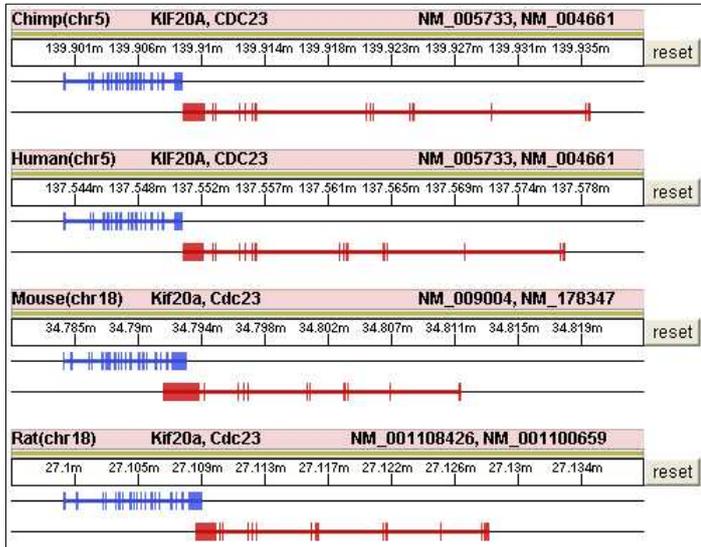


그림 8: 진화분석을 위한 종별 특정 두 유전자 비교

### 3.4 Ruler 클래스

Ruler 클래스는 검색된 유전자의 위치를 나타내고 마우스 드래깅으로 부분 확대하는 기능을 지원한다. 그림 9에 검색된 유전자의 부분 확대 기능을 보이고 있다. 확대된 후 상단부의 작은 막대바는 특정 유전자 검색결과 전체 범위 중 확대된 현재 위치를 나타내며 reset버튼을 눌러 원래대로 돌아갈 수 있다.

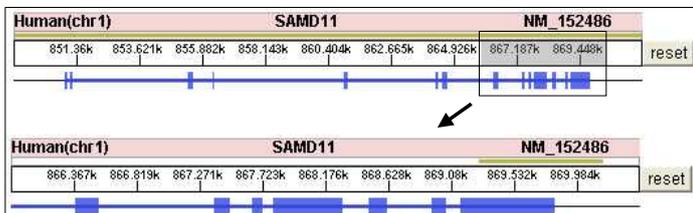


그림 9: 검색된 유전자의 부분 확대

### 3.5 Query Agent 클래스

Query Agent 클래스는 클라이언트의 Query Manager 클래스의 쿼리를 받아 DBMS와 연동하여 쿼리를 검색하는 서버측 클래스이다. 유전자 검색 기능 중 가장 수행 시간이 많이 걸리는 부분이며, 이 기능을 상대적으로 높

은 성능의 서버측에서 수행한다. 이 Query Agent 클래스만이 DBMS와 직접 연동하게 되며 검색된 결과는 클라이언트의 Query Manager로 전송된다.

## 4. Component Ware의 클라이언트

본 논문에서 제안하는 시스템은 Java로 구현되며, 클라이언트 부분은 Java Applet으로 구현된다. Java언어의 플랫폼 독립적 특성 때문에 다양한 시스템 환경을 위한 어플리케이션 제작에 Java언어가 좋은 도구가 된다. 그러나 Java Applet의 경우 접근하는 클라이언트의 종류에 따라 약간의 문제점이 발생된다. 클라이언트의 웹 브라우저에 따라 JRE (Java Runtime Environment)가 클릭 몇 번으로 자동설치가 될 수도 있고, 수동으로 설치를 해야할 수도 있기 때문이다. 그림 10은 JRE가 설치되지 않은 클라이언트가 접속했을 때 자동으로 JRE가 설치되기 전의 화면이다.

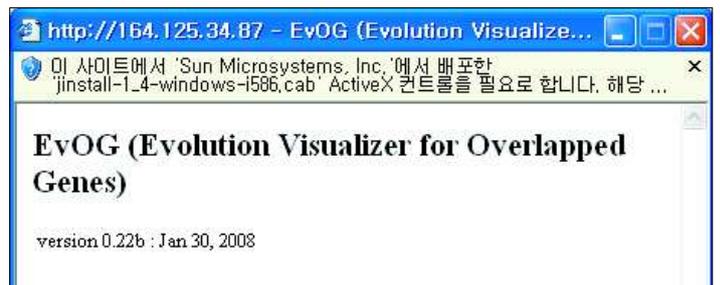


그림 10: JRE 자동설치 대기화면

상단부를 클릭했을 때 JRE환경이 Sun Microsystems사의 서버에서 다운로드되어 설치된다. 클라이언트측의 사용자는 세부내역을 알지 못하더라도 쉽게 접근할 수 있다. 그림 10의 수행을 위해서는 기존의 Applet코드를 연결하는 HTML문서에 Java의 HTMLConverter 명령을 적용하면 된다. 그림 11은 HTMLConverter를 수행했을 때의 화면이다. 이 명령을 적용하면 HTML 코드내에 JRE 자동설치 부분이 추가된다. Java Applet 기반의 S/W들은 JRE 설치 여부를 확인하고 JRE가 설치되지 않은 경우 정상적으로 서비스를 사용할 수 없다는 내용을 표시하게 된다. 현재 Microsoft의 Internet Explorer는 Sun사에서 자동적으로 JRE가 설치될 수 있도록 기능을 지원하고 있다.



그림 11: HTMLConverter 실행화면

## 5. 실험결과

본 논문에서는 유전자 진화분석을 위한 특정 유전자 정보를 종별로 비교할 수 있는 유전자 가시화 도구 개발용 Component Ware를 제안하였다. 이 Component Ware는 개별 유전자를 종(species)과 염색체 등의 부가 정보와 함께 가시화함으로써 진화분석을 위한 용도뿐 아니라 자신의 연구목적에 맞는 S/W를 쉽게 제작하는데 사용될 수 있었다. 또한 두 유전자의 관계 분석은 진화 분석과 함께 두 유전자의 상호 관계를 파악할 수 있는 다른 생물학적 연구에도 도움을 줄 수 있었다. 본 논문에서 제안한 Component Ware로 생성된 Gene Browser는 S/W를 업그레이드 하더라도 다음 번 접속시에 자동적으로 코드가 다운로드 되어 실행된다. 또한 많은 실행 시간이 필요한 코드는 서버측에서 실행함으로써 수행속도와 S/W의 유지보수에도 장점이 있었다.

## 6. 참고문헌

[1] Chris Mayor, Michael Brudno, Jody R. Schwartz, Alexander Poliakov, Edward M. Rubin, Kelly A. Frazer, Lior S. Pachter, and Inna Dubchak. "VISTA : visualizing global DNA sequence alignments of arbitrary length," *Bioinformatics*, 16(11), pp 1046-1047, 2000.

[2] Ivan Ovcharenko, Marcelo A. Nobrega, Gabriela G. Loots and Lisa Stubbs. "ECR Browser: a tool for visualizing and accessing data from

comparisons of multiple vertebrate genomes," *Nucleic Acids Research*, 32, 2004.

[3] Scott Schwartz, Zheng Zhang, Kelly A. Frazer, Arian Smit, Cathy Riemer, John Bouck, Richard Gibbs, Ross Hardison, and Webb Miller. "PipMaker—A Web Server for Aligning Two Genomic DNA Sequences," *Genome Research*, 10, pp 577-586, 2000.

[4] Reinhard Engels, Tamara Yu, Chris Burge, Jill P. Mesirov, David DeCaprio and James E. Galagan. "Combo: a whole genome comparative browser," *Bioinformatics*, 22(14), pp 1782-1783, 2006.

[5] Karolchik, D. et al. 2008. The UCSC Genome Browser Database: 2008 update, *Nucleic Acids Research*, Vol. 36, pp. D773-D779.

[6] Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell*. 1993 Dec 3;75(5):855-62.

[7] Dallosso AR, Hancock AL, Malik S, Salpekar A, King-Underwood L, Pritchard-Jones K, Peters J, Moorwood K, Ward A, Malik KT, Brown KW. Alternately spliced WT1 antisense transcripts interact with WT1 sense RNA and show epigenetic and splicing defects in cancer RNA. 2007 Dec;13(12):2287-99. Epub 2007 Oct 16.