

K-Viz: 대사 경로 비교를 위한 KEGG 기반의 시각화

(K-Viz: KEGG Based Bisualization for Comparing Metabolic Pathways)

임 동 혁 [†] 이 동 희 [†] 김 형 주 ^{**}

(Dong-Hyuk Im) (Dong-Hee Lee) (Hyoung-Joo Kim)

요약 서로 다른 종에서의 대사 경로를 비교하는 것은 아직 밝혀지지 않은 유전자를 찾는 문제에 있어 매우 중요하다. 많은 대사 경로 시각화가 제시되었지만 생물학자들은 대사 경로를 단순히 보는 것뿐만 아니라 비교하는 시각화를 필요로 한다. K-Viz는 KEGG 기반의 대사 경로를 위한 시각화 툴이다. 다른 종의 대사 경로를 비교하기 위해 KEGG의 PathComp와 같은 경로를 서로 다른 색으로 표현하는 방법과 각 종에서의 경로를 테이블로 보여준다. K-Viz는 생물학자들이 서로 다른 종에서의 대사 경로 비교를 이해하는데 도움을 준다.

키워드 : 시각화, KEGG, 대사 경로

Abstract The comparison of metabolic pathway in different species is important in detecting a missing gene. There are many visualizations for metabolic pathway. However, Biologists need not only a simple path but also a visualization for comparison. K-Viz is a tool for visualization of metabolic pathway based on KEGG. To compare pathways in different species, K-Viz uses different color for path such as PathComp in KEGG and shows the table of path in pathway. K-Viz helps biologists to understand the comparison of metabolic pathways in different species.

Key words : visualization, KEGG, metabolic pathway

1. 서론

대사 경로(metabolic pathway)란 생명체 내에서의 화합물에서의 화학 반응을 나타내준다. 경로에 관여하는 화합물과 이러한 화합물을 변환시켜주는 효소(Enzyme)에 대한 정보 즉 효소에 해당되는 유전자의 특징과 대사 경로 상에서의 위치를 나타내준다. 이러한 대사 경로는 물질 처리 과정 등의 분석과 비교를 하는데 있어 중요성을 가지게 된다. 특히 생물학자들은 대사 경로 분석을 시각화하여 사용하는 것이 효과적이다. 이러한 대사 경로 시각화는 대사 경로에서의 다양한 처리 과정에 연

관 있는 효소의 경로와 화합물에 대한 다이어그램 정보를 표현한다. 대사 경로 시각화는 크게 정적 시각화와 동적 시각화로 구분된다. 정적 시각화(static visualization)는 이미 대사 경로 지도를 만들어 보여주는 것으로 KEGG가 가장 대표적인 경우이다[1,2]. 이런 정적 시각화는 한 번 만들어진 지도를 계속해서 사용하기 때문에 자주 사용되는 경우에 적합하다. KEGG의 대사 경로 지도는 현재 가장 많이 사용되고 있다. 동적 시각화(dynamic visualization)는 정적 시각화와는 반대로 사용자가 요구할 때 대사 경로 지도를 그려 보여준다. 이러한 동적 시각화의 장점은 정적 시각화의 유지 비용보다 훨씬 적게 든다는 것이다. 동적 시각화의 대표적인 시스템으로는 PathFinder[3], PaVESy[4], Pathway Database System[5]등이 있다. 하지만 정적 시각화나 동적 시각화는 단순히 대사 경로를 보여주는 것으로 대사 경로를 비교하는 데는 한계를 가지고 있다. 이런 한계를 극복하기 위한 시각화 방법으로 대사 경로를 다른 색으로 표현하는 방법이 이용되고 있다[6]. 각각의 종마다 다른 색으로 표시를 하여 대사 경로를 비교할 수 있

· 본 연구는 BK-21 정보기술 사업단과 정보통신부 및 정보통신연구진흥원의 대학 IT연구센터 육성지원사업(ITA-2005-C1090-0502-0016)의 연구 결과로 수행되었음

[†] 학생회원 : 서울대학교 컴퓨터공학
dhim@idb.snu.ac.kr
leedh@idb.snu.ac.kr

^{**} 종신회원 : 서울대학교 컴퓨터공학 교수
hjk@snu.ac.kr

논문접수 : 2006년 11월 2일

심사완료 : 2007년 3월 24일

게 해준다. 이런 종마다 다른 색으로 표시하는 시각화의 단점은 각 종마다 다른 색으로 표시를 하기 때문에 종 이 많아지면 알아보기 어렵다는 단점이 있다.

본 논문에서는 제한한 K-Viz는 KEGG 대사 경로 지도를 기반으로 하는 대사 경로 비교를 위한 시각화이다. KEGG 대사 경로 지도 위에 여러 종의 대사 경로 지도를 겹쳐서 해당되는 효소의 횟수를 세어서 그 횟수 별로 색을 다르게 표시한다. 또한 추가적인 정보로 대사 경로를 각 종별로 테이블과 같이 보여주어 대사 경로 지도 비교를 쉽게 이해 할 수 있게 해준다. 이런 대사 경로 지도 비교는 발견되지 않은 유전자를 찾는 데 유용하게 사용될 수 있다[7]. 비슷한 종에서 대사 경로 지도에서의 유전자를 비교하여 각각의 종마다 대사 경로 지도에서의 특정 유전자의 존재 유무를 파악하여 생물학자들이 이 유전자에 대한 연구를 도와줄 수 있다. 본 논문의 구성은 다음과 같다. 2장에서는 관련 연구로서 대사 경로에서 사용되는 여러 가지 시각화 툴을 알아보고 3장에서는 K-Viz를 설명하고 4장은 타 대사 경로 시각화의 비교를 보여주며 마지막으로 결론으로 구성되어 있다.

2. 관련 연구

KEGG[2]는 가장 대표적인 대사 경로 시스템이다. KEGG(Kyoto Encyclopedia of Genes and Genomes)

는 교토 대학에서 제공하는 GenomeNet의 핵심부분으로 KEGG 데이터 베이스는 생명체의 다양한 처리와 연관 있는 효소와 화합물에 대한 다이어그램을 포함하고 있다. 데이터의 관계를 그래프로 표현하여 사용한 특징을 가지고 있다. 그림 1에서와 같이 KEGG 대사 경로 지도는 전체적인 대사 경로를 표현한다. 사각형이 화학 반응을 일으키는 효소를 나타내며 동그라미는 화합물을 나타낸다. 특정 종에 대한 대사 경로는 녹색으로 효소를 표현하여 보여준다.

PathFinder[3]는 대사 경로 시각화에서 가장 동적인 시각화를 보여준다. 처음과 끝 위치를 지정해주면 그 사이의 대사 경로를 동적으로 보여준다. 단순히 경로를 보여주는 것 이외에 이 경로가 타당한 경로인지 부분경로(subway)와 정크(chunk)를 두어 확인을 한다. 각 부분경로는 갈라짐이 없는 단위 경로인 정크로 나누어진다. 따라서 각 부분경로는 일련의 정크로 구성되어 있다. 여기서 각 정크에 나타나는 효소 중에 해당 종에 있는 효소의 비율을 정크 점수라 하고 사용자가 정한 임계값보다 큰 점수를 가진 것들로만 이루어진 부분경로가 있을 경우 그것을 유효한 부분경로(valid subway)라 한다. 그 경로가 있을 가능성이 높으므로 이 경로상에 나타나 효소 중에 없는 것을 더 찾아보라고 사용자에게 권하게 되는 것이다. 그림 2는 PathFinder에서의 경로를 보여 준다.

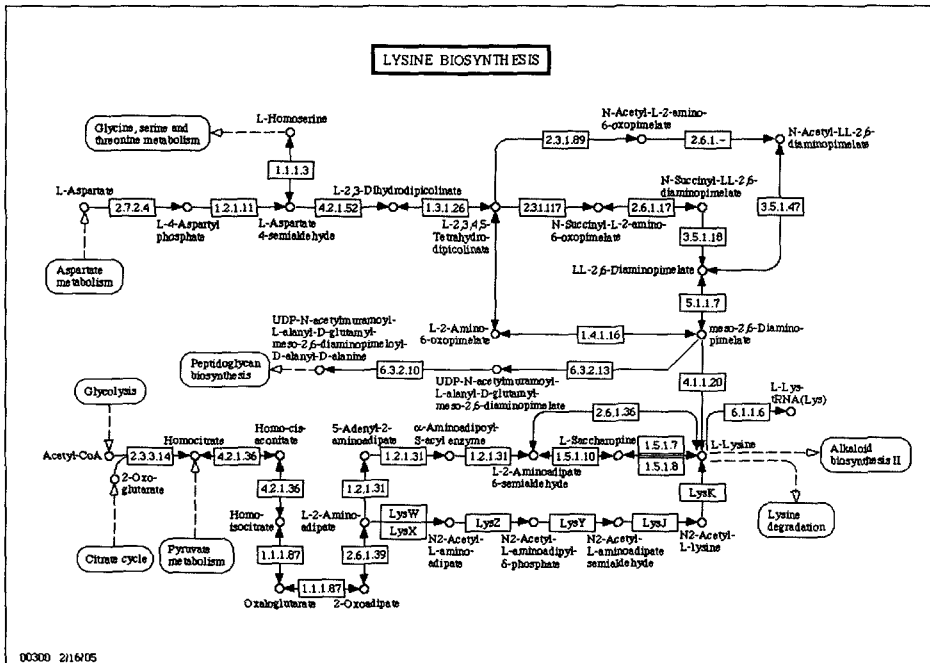


그림 1 KEGG에서의 대사 경로 지도 시각화

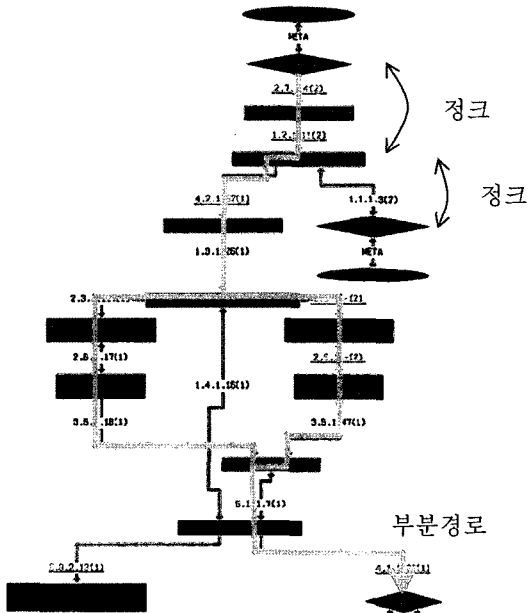


그림 2 PathFinder에서의 대사 경로

K-Viz가 KEGG 대사 경로 시각화와 PathFinder와의 가장 큰 차이점은 K-Viz는 단순한 대사 경로를 시각화할 뿐 아니라 여러 종에 대한 대사 경로를 비교함으로써 발견되지 않은 효소를 찾아 줄 수 있는 기능을 제공하고 있다는 것이다. PathFinder는 대사 경로 비교보다는 하나의 종에 대해 대사 경로의 타당성을 제공해 줄 뿐이고 KEGG 대사 경로 시각화에서는 종에 대한 비교를 제공하지만 종의 개수가 제한되며 비교도 대사 경로가 아닌 지정되는 효소에 대한 비교를 시각화 해준다. 이에 반해 K-Viz에서는 17가지 미생물을 대상으로 하며 대사 경로 전체 혹은 일정 대사 경로에 대한 비교를 시각화로 제공해준다.

대사 경로에서 여러 종에 나타나는 공통되는 부분을 찾아내려는 접근이 있다[8]. 이러한 접근은 대사 경로를 하나의 그래프로 간주하여 공통되는 최대 부분그래프를

찾아내는 것이다. 최대 부분그래프 마이닝을 위한 깊이 우선 탐색 알고리즘을 제시하였으며 임계값을 변경해가며 흥미로운 결과를 나타내었다. 각 노드들의 나타나는 횟수를 두어 임계값 이상의 노드들로 구성된 최대 부분그래프에서 의미 있는 패턴을 찾으려고 한다. K-Viz와 횟수를 두어 경로를 표시한다는 비슷한 점이 있지만 이 논문의 경우는 경로가 아니라 부분그래프 즉 공통된 패턴을 찾는다는 것이 가장 큰 차이일 수 있다.

3. K-Viz

K-Viz는 대사 경로 비교를 위한 시각화를 제공한다. 이외에도 대사 경로 분석을 위한 기능 또한 포함하고 있다.

3.1 대사 경로 모델

대사 경로는 그래프로 표현된다. 하지만 본 논문에서의 대사 경로는 방향이분그래프(directed bipartite graph)로 표현을 한다. $G = (V_1 \cup V_2, E)$ 이다. 노드 $v \in V_1$ 로서 이 노드는 화합물(compound)의 집합을 표현하며 노드 $v \in V_2$ 는 효소(Enzyme), 화학 반응(Reaction)의 집합을 나타낸다. 대사 경로에서의 각각의 구성은 이러한 노드들과 노드들의 관계로 시각화 되어진다. 이러한 모델을 표현하기 위해 K-Viz에서 사용되는 데이터 모델은 RDF 데이터 모델을 사용한다. RDF 모델을 사용하는 이유는 이러한 대사 경로의 데이터 모델을 잘 표현할 뿐만이 아니라 시각화에서는 대사 경로 지도에서의 노드들 간의 관계 및 지도에서의 위치, 형태 등의 메타 정보만 필요하다. 그림 3에서는 이러한 RDF 모델 중의 효소에 대한 예를 보여주고 있다.

그림 3에서 효소의 이름은 EC#1.1.1.2를 나타내며 대사 경로의 지도에서 가로 45, 높이 17인 사각형의 형태를 지니며 좌표 값이 (310,977)에 시각화 되어진다. 그림 3과 같은 방법으로 화합물도 표현이 된다.

3.2 대사 경로 질의

대사 경로에 사용되는 질의는 그래프 모델에서의 경로 질의와 같으며 다음과 같은 정의로 표현할 수 있다.

```
<k:entry>
  <Enzyme rdf:nodeID="_14">
    <k:name rdf:resource="http://www.w3.org/2005/02/13-KEGG/ec#1.1.1.2"/>
    <k:graphics>
      <Rectangle k:name="1.1.1.2" k:fgcolor="#000000" k:bgcolor="#FFFFFF"
        k:x="310" k:y="977" k:width="45" k:height="17"/>
    </k:graphics>
  </Enzyme>
</k:entry>
```

그림 3 대사 경로 효소에 대한 RDF 데이터

정의 1. 대사 경로 $P(C, E, R)$ 는 화합물 C , 화학 반응 R , 효소 E 로 구성된다. $r \in R$ 인 화학 반응은 기질 (substrate) 집합 $S(r) \subseteq C$ 과 생성 물질 집합 $T(r) \subseteq C$ 의 조합과 효소 E 으로 구성된다.

위 정의는 하나의 대사 경로에서의 화학 반응은 기질과 생성물로 이루어진다. 구체적으로 하나의 화합물이 효소에 의해 다른 화합물 즉 생성물이 되는 것이 화학 반응을 구성하게 되는 것이다. 이런 화합물은 기질이 될 수도 있고 생성물이 될 수도 있다.

정의 2. 대사 경로에서의 경로 $Path$ 는 대사 경로 P 에서의 $1 < i \leq N$ (N 은 임계값)이고 $T(r_i) = S(r_{i+1})$ 을 만족하는 화학 반응 $r_1, r_2, r_1 \dots r_n$ 의 연속이다.

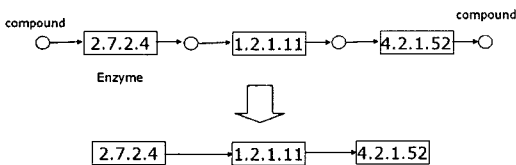


그림 4 대사 경로 비교를 위한 화학 반응 리스트

정의 2는 대사 경로에서 사용되는 경로를 정의하고 있다. 어떤 특정한 임계값을 넘지 않고 이전 화학 반응의 생성물과 다음 기질이 같은 모든 화학 반응의 연속적인 리스트를 경로라 정의한다. 그림 4에서 보여주듯이 원래의 전체 경로가 화학 반응만의 연속적인 리스트로 바뀌어지는 것으로 대사 경로에서 화합물이 제거되는 것을 의미한다. 위와 같은 경로 질의는 KEGG의 Path-Comp[9] 등에서 직접 사용되고 있으며 K-Viz에서도 사용을 하고 있다.

정의 3. 각각의 대사 경로의 경로 $Path_1, Path_2, \dots, Path_n$ 는 중 $1, 2, \dots, n$ 에서의 경로를 나타낸다.

정의 3은 대사 경로를 비교할 대상을 정의하고 있다. 구체적으로 K-Viz의 목표는 $Path_1, Path_2, \dots, Path_n$ 을 비교하여 시각화로 보여주는 것이며 각각의 종이 가지는 화학 반응 리스트를 비교하는 것이다.

K-Viz에서의 질의는 위의 정의에 따른다. 또한 대사 경로 데이터가 RDF로 되어 있기 때문에 RDF 질의처 리에서 가장 많이 사용되는 SeRQL[10]을 이용한다. SeRQL은 Sesame[11]에서 제공해주고 있다. 그림 5는 SeRQL을 이용하여 대사 경로에서의 $Path_i$ 를 구성하는 화학 반응 r_i 의 위치를 질의하는 예를 보여주고 있다.

그림 5의 SeRQL로 얻어진 결과를 바탕으로 대사 경로를 구성하는 각 구성 요소의 위치와 형태를 언어서 지도에 시각화를 보여줄 수 있다.

3.3 시스템 구현

K-Viz는 JAVA로 구현되었으며 웹 기반의 시각화 틀

```
Select x, y, width, height
from {Pathway} k:name {mapID},
    {Pathway} k:entry {_bEntry},
    {_bEntry} k:reaction { r_i },
    {_bEntry} k:graphics {Graphics},
    {Graphics} k:x {x},
    {Graphics} k:y {y},
    {Graphics} k:width {width},
    {Graphics} k:height {height}
using namespace rdf=<http://www.w3.org/1999/02/22-rdf-syntax-ns#>,
    k=<http://www.w3.org/2005/02/13-KEGG/#>,
    path=<http://www.w3.org/2005/02/13-KEGG/path#>,
    rn=<http://www.w3.org/2005/02/13-KEGG/rn#>
```

그림 5 SeRQL을 이용한 대사 경로 질의

이다. 대사 경로 데이터는 KEGG의 KGML[12]을 [13]에서의 매핑 방법으로 RDF로 변환하여 사용하였다. RDF 저장소로는 Sesame를 이용하였다. K-Viz에서 다루는 좋은 미생물 17종이다. 미생물 대사 경로에 대한 연구는 매우 활발히 진행 중이다. 미생물에서의 대사 경로는 분석하기가 쉽고 또한 실험으로 밝혀진 것이 많은 장점을 가진다. 게다가 대사 경로가 많이 밝혀진 미생물 종끼리의 대사 경로 비교를 통해 발견되지 않은 유전자를 찾는 것은 매우 의미 있는 일이다. 표 1은 K-Viz에서 사용한 17가지 미생물을 보여주고 있다.

표 1 K-Viz에서 사용하는 미생물 17 종

K-Viz에서 사용되는 미생물 중
Haemophilus influenza, Escherichia coli K12, Xylella fastidiosa 9a5c, Vibrio cholerae, Pseudomonas aeruginosa, Pasteurella multocida, Pseudomonas putida, Salmonella typhimurium LT2, Salmonella Typhi CT18, Xanthomonas axonopodis, Yersinia Pestis KM, Shewanella oneidensis, Vibrio vulnificus CMCP6, Buchnera aphidicola str. Bp, Pseudomonas syringae DC3000, Vibrio parahaemolyticus, Shigella flexneri 2a str.

K-Viz는 크게 3가지의 시각화를 제공해주고 있다. 첫 번째로 특정 유전자의 대사 경로를 보여준다. 예를 들면 Glycolysis/ Gluconeogenesis 에서의 효소 EC:6.2.1.1에 해당되는 b4069 유전자(acs,yfaC, acsA)의 대사 경로에서의 시각화를 그림 6처럼 보여준다.

그림 6에서 알 수 있듯이 사용자는 해당 유전자의 대사 경로에서의 위치를 한 눈에 알 수 있다. 두 번째 기능으로서 각 종에서의 대사 경로에 나타나는 모든 유전자들을 비교할 수 있도록 지도를 겹쳐서 횡수를 두어 유전자가 속하는 효소의 색을 다르게 표현함으로써 각 종에서의 유전자 유무를 알 수 있게 해준다. 효소 테이블을 추가하여 시각화보다 상세한 정보를 같이 제공해주는 장점을 가진다.

그림 7은 Glycolysis / Gluconeogenesis에서의 17가

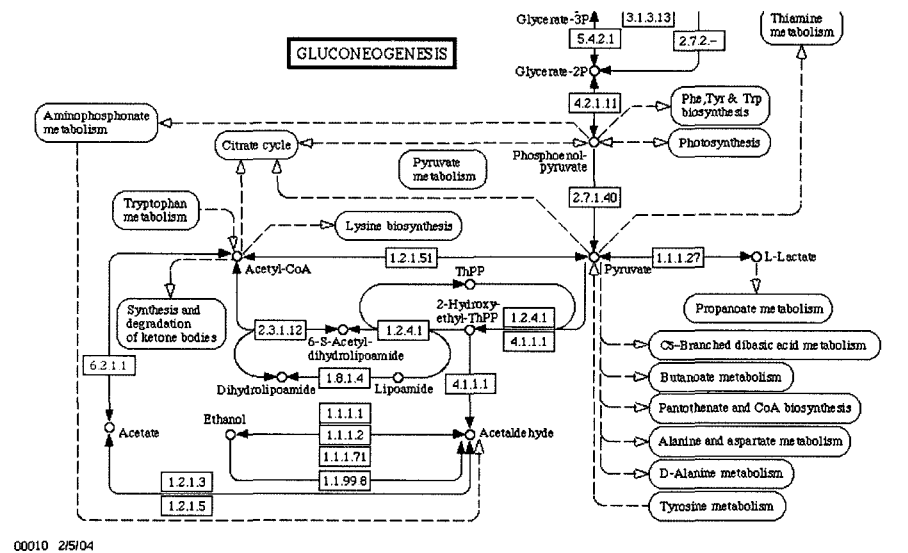


그림 6 특정 유전자의 대사 경로 분석을 위한 시각화

지 미생물에 대한 밝혀진 유전자에 대한 시각화를 보여 주고 있다. 그림 7의 (a)에서 알 수 있듯이 Glycolysis / Gluconeogenesis의 지도에 나타난 색을 가지고 모든 종에 나타나는 효소나 그 효소에 속한 유전자를 알 수 있다. 또한 그림 7의 (b)를 통해 종별로 구분할 수 있으며 효소 EC:2.7.1.12는 모든 종에서 나타나는 것을 알 수 있으며 효소 EC:2.7.1.69는 오직 *eco(E. coli)*에서만 나타나는 것을 알 수 있다. K-Viz에서는 유전자에 대해 상세한 정보를 포함하지 않고 있다. RDF로 메타 정보 만으로 표현이 되기 때문에 좀더 자세한 정보를 필요로 하는 경우를 위해 KEGG의 DBGET[14]으로 링크를 두어 정보를 제공해준다. 그림 8은 이러한 링크 정보를 보여주고 있다.

그림 8과 같이 특정 효소를 클릭하면 해당 종에서의 유전자에 대한 상세정보를 선택하여 얻을 수 있다.

K-Viz가 제공하는 세 번째 기능은 여러 종에서의 대사 경로 비교를 제공하는 것이다. KEGG의 PathComp는 대사 회로 내에서의 경로를 예측하는 응용 프로그램이다. PathComp는 또한 결과를 시각화 하여 보여주는 기능을 가지고 있으나 단순히 경로를 보여주지만 한다. 그림 9는 KEGG의 PathComp에서 제공해주는 시각화이다. 그림 9의 경로는 Lysine-biosynthesis에서의 L-Aspartate에서 LL-2, 6-Diaminopimelate까지의 경로로 빨간색으로 표시된다. 하지만 다른 종과의 비교를 위해서는 같은 방법으로 다른 종에서도 PathComp를 이용하여 경로를 구해야 하며 종이 추가되면 비교가 더 어려워지는 제한을 가지게 된다.

이러한 단점을 극복하기 위해 K-Viz는 대사 경로에

서의 경로를 비교할 수 있는 기능을 제공해준다. 경로에 대한 시각화 방법은 두 번째 기능과 같으나 효소 테이블을 각 종에 대한 경로로 보여주는 것에 차이가 있다. 이 기능은 3.2절에서 정의한 질의이며 사용자가 원하는 경로만을 여러 종에 겹쳐서 표현된다. 그림 10은 K-Viz에서의 대사 경로 비교를 위한 시각화를 보여준다. 경로는 그림 9와 같으며 종은 *eco*, *ypk*, *ych*, *vch*, *stm*의 5가지를 비교하였다. 그림 10의 (a)에서 알 수 있듯이 정해진 경로만이 색으로 표현되며 각 효소에서 횟수를 두어 횟수에 맞는 색으로 표현된다. 따라서 그림에서 EC:2.6.1.17의 색이 다른 효소들과 다른 것을 알 수 있으며 이 효소가 발견되지 않았다는 것을 알 수 있다. 그림 10의 (b)는 경로별 효소의 존재 유무를 보여주는 테이블이다. 테이블을 보면 *vch*와 *vpa*종에 대한 경로가 다른 종에 비해 불완전한 것을 알 수 있다.

3.4 시스템 비교

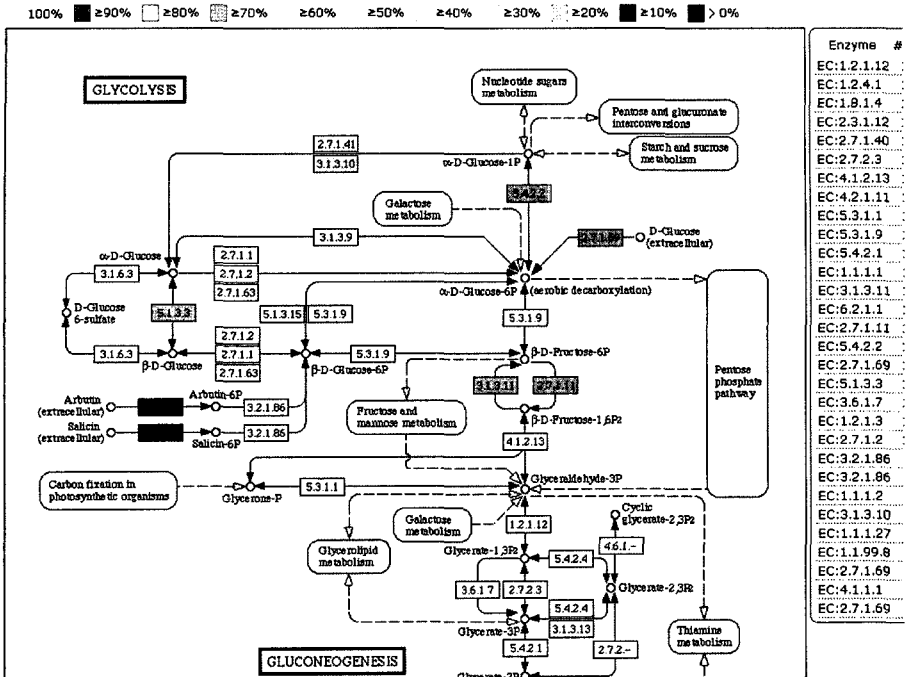
대사 경로 시각화에 대한 대표적인 시스템과 K-Viz를 비교하였다. 비교 대상으로는 KEGG, PathFinder로 각각 정적 시각화와 동적 시각화를 대표하는 시각화 툴로써 포함이 되었다. 표 2는 K-Viz와 타 시스템과의 비교를 보여주고 있다.

표 2에서 알 수 있듯이 K-Viz의 장점은 대사 경로 분석이 가능하며 또한 이러한 경로를 여러 종에서 비교가 가능하다는 것이다.

4. 결론

지금까지 K-Viz의 기능과 장점에 대해 알아보았다. K-Viz는 KEGG 기반의 대사 경로 지도에 색으로 표현

Metabolism > Glycolysis / Gluconeogenesis



(a) K-Viz에서의 여러 종에서 밝혀진 효소들의 시각화

Enzyme	#/17	bab	eco	hin	pae	pmu	ppu	pst	sfl	son	stm	stt	vch	vpa	vvu	xac	xfa	ypk
EC:1.2.1.12	17	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:1.2.4.1	17	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:1.8.1.4	17	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:2.3.1.12	17	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:2.7.1.40	17	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:2.7.2.3	17	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:4.1.2.13	17	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:4.2.1.11	17	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:5.3.1.1	17	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:5.3.1.9	17	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:5.4.2.1	17	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:1.1.1.1	16	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:3.1.3.11	15	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:6.2.1.1	14	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:2.7.1.11	13	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:5.4.2.2	13	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:2.7.1.69	12	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:5.1.3.3	12	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:3.6.1.7	11	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:1.2.1.3	10	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:2.7.1.2	10	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:3.2.1.86	9	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:3.2.1.86	7	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:1.1.1.2	5	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:3.1.3.10	4	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:1.1.1.27	2	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:1.1.99.8	2	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:2.7.1.69	2	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:4.1.1.1	2	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
EC:2.7.1.69	1	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o

(b) 대사 경로에 나타나는 효소들의 존재 유무를 보여주는 효소 테이블
그림 7 K-Viz에서의 효소 시각화

참고 문헌

- [1] Hiroyuki Ogata, Susumu Goto, Kazushige Sato, Wataru Fujibuchi, Hidemasa Bono, "KEGG: Kyoto Encyclopedia of Genes and Genomes," Nucleic Acids Research, 2000.
- [2] KEGG, <http://www.genome.ad.jp/kegg/>
- [3] Alexander Goesmann, Martin Haubrock, Folker Meyer, Jorn Kalinowski, Robert Giegerich, "Path-Finder: reconstruction and dynamic visualization of metabolic pathways," Bioinformatics, vol. 18, 2002.
- [4] Alexander Ludemann, Daniel Weicht, Joachim Selbig, Joachim Kopka, "PaVESy: Pathway Visualization and Editing System," Bioinformatics, vol. 20, 2004.
- [5] L. Krishnamurthy, J. Nadeau, G. Ozsoyoglu, M. Ozsoyoglu, G. Schaeffer, M. Tasan, W.Xu, "Pathway Database System: An integrated set of tools for biological pathways," Bioinformatics, vol. 19, 2003.
- [6] Forster M., Pick A., Raitner M., Schreiber F., Brandenburg F. J., "The System Architecture of the BioPath System," In Silico Biology, 2002.
- [7] Andrei Osterman, Ross Overbeek, "Missing genes in metabolic pathways: a comparative genomics approach," Current Opinion in Chemical Biology, 2003.
- [8] Mehmet Koyuturk, Ananth Grama, Wojciech Szpankowski, "An efficient algorithm for detecting frequent subgraphs in biological networks," Bioinformatics, vol. 20, 2004.

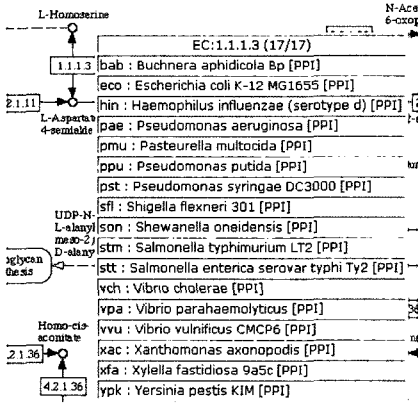


그림 8 효소에 속하는 유전자의 DBGET 링크

하는 기존의 비교 방법을 대사 경로의 특정한 경로에 적용을 하여 시각화로 보여준다. 그 결과 여러 종에서의 특정한 경로에 대한 비교를 알아보기 쉽게 할 수 있다. 이런 특정 경로에 대한 비교가 가지는 의미는 발견되지 않은 유전자를 쉽게 찾을 수 있다는 것이다. 향후 연구로는 찾아진 유전자에 대한 분석이 필요하다. 현재 발견되지 않은 유전자에 대한 연구로 Genome Context Analysis가 활발하게 진행 중이며 발견되지 않은 유전자에 대해 올바른 유전자를 제공해주는 것이 필요하다. 유전자 클러스터 방법, occurrence profile 등의 방법이 많이 사용되고 있다. 향후 이러한 방법과 연계하여 시각화하면 좀 더 효과적일 것이다.

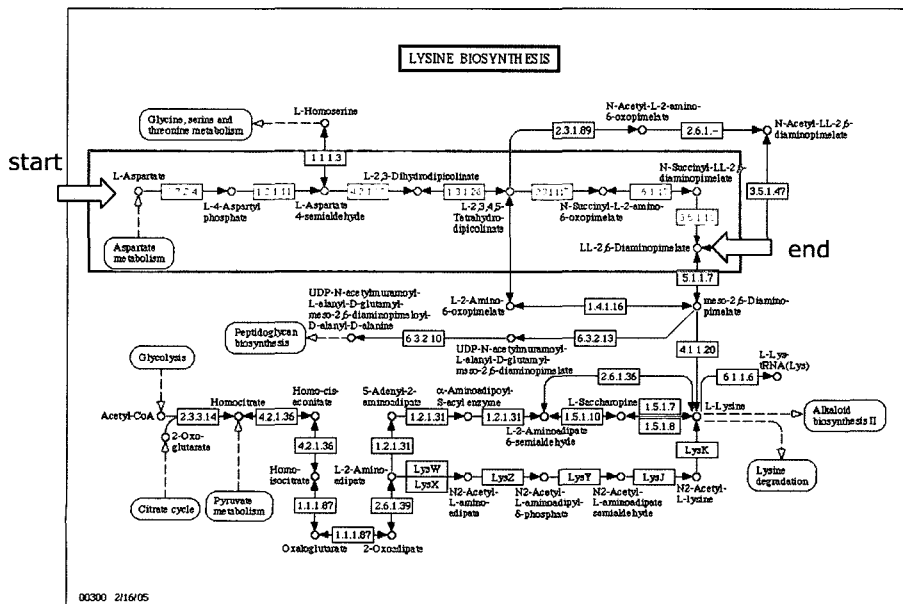
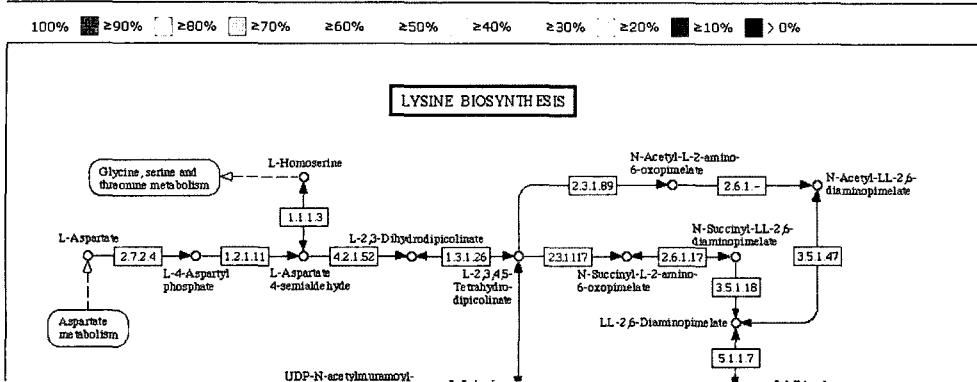
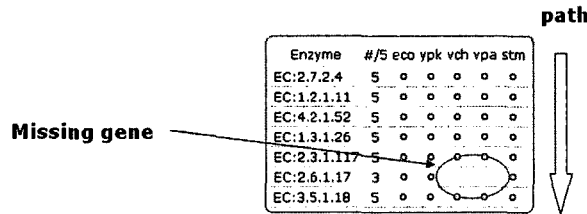


그림 9 KEGG PathComp에서의 대사 경로 시각화

Metabolism > Lysine biosynthesis



(a) K-Viz에서의 종 비교를 위한 경로 시각화



(b) K-Viz에서의 경로 비교 테이블

그림 10 K-Viz의 시각화

표 2 K-Viz 시스템 비교

	KEGG	PathFinder	K-Viz
시각화 방법	정적 + 색으로 표현	동적	정적 + 색으로 표현
대상 종	모든 종	모든 종	미생물 17종
대사 경로 분석	가능	불가능	가능
경로 비교	특정 효소들만 가능	비교 불가	경로 비교 가능

[9] PathComp, www.genome.jp/kegg/tool/pathcomp.html
 [10] Jeen Broekstra, Arjohn Kampman, "SeRQL: An RDF Query and Transformation Language," <http://openrdf.org>, 2004.
 [11] J. Broekstra, A. Kampman, F. Harmelen, "Sesame: An Architecture for Storing and Querying RDF Data and Schema Information," International Semantic Web Conference 2002.
 [12] KGML, <http://www.genome.jp/kegg/xml/>
 [13] KEGG RDF Mapping, <http://www.w3.org/2005/02/13-KEGG/>
 [14] DBGET Search, www.genome.jp/dbget/



임동혁
 2003년 고려대학교 컴퓨터교육과(학사)
 2005년 서울대학교 컴퓨터공학부(석사)
 2005년~현재 서울대학교 컴퓨터공학부 박사과정 재학 중. 관심분야는 데이터베이스, 시맨틱 웹, 바이오인포메틱스



이동희
 2005년 경북대학교 컴퓨터공학과(학사)
 2007년 서울대학교 전기컴퓨터공학부(석사). 2007년~현재 한국오라클 재직 중. 관심분야는 데이터베이스, 시맨틱 웹, 바이오인포메틱스



김형주
 1982년 서울대학교 전산학과(학사). 1985년 Univ. of Texas at Austin(석사)
 1988년 Univ. of Texas at Austin(박사). 1998년 5월~1988년 9월 Univ. of Texas at Austin(Post-Doc). 1988년 9월~1990년 12월 Georgia Institute of Technology(부교수). 1991년~현재 서울대학교 컴퓨터공학부 교수. 관심분야는 데이터베이스, XML, 시맨틱 웹, 온톨로지