

## 닭 특이 대사 경로 재확립

김운수<sup>1</sup> · 이세영<sup>1</sup> · 박혜선<sup>1,2</sup> · 백운기<sup>3</sup> · 이준현<sup>4</sup> · 서성원<sup>1,\*</sup>

<sup>1</sup>충남대학교 동물바이오시스템학과, <sup>2</sup>농촌진흥청 국립축산과학원, <sup>3</sup>국립중앙과학관, <sup>4</sup>충남대학교 동물자원생명과학과

### Reconstruction of Metabolic Pathway for the Chicken Genome

Woon Su Kim<sup>1</sup>, Se Young Lee<sup>1</sup>, Hye Sun Park<sup>1,2</sup>, Woon Kee Baik<sup>3</sup>, Jun Heon Lee<sup>4</sup> and Seongwon Seo<sup>1,\*</sup>

<sup>1</sup>Department of Animal Biosystem Science, Chungnam National University, Daejeon 305-764, Korea

<sup>2</sup>National Institute of Animal Science, RDA

<sup>3</sup>National Science Museum

<sup>4</sup>Department of Animal Science and Biotechnology, Chungnam National University, Daejeon 305-764, Korea

**ABSTRACT** Chicken is an important livestock as a valuable biomedical model as well as food for human, and there is a strong rationale for improving our understanding on metabolism and physiology of this organism. The first draft of chicken genome assembly was released in 2004, which enables elaboration on the linkage between genetic and metabolic traits of chicken. The objectives of this study were thus to reconstruct metabolic pathway of the chicken genome and to construct a chicken specific pathway genome database (PGDB). We developed a comprehensive genome database for chicken by integrating all the known annotations for chicken genes and proteins using a pipeline written in Perl. Based on the comprehensive genome annotations, metabolic pathways of the chicken genome were reconstructed using the PathoLogic algorithm in Pathway Tools software. We identified a total of 212 metabolic pathways, 2,709 enzymes, 71 transporters, 1,698 enzymatic reactions, 8 transport reactions, and 1,360 compounds in the current chicken genome build, Gallus\_gallus-2.1. Comparative metabolic analysis with the human, mouse and cattle genomes revealed that core metabolic pathways are highly conserved in the chicken genome. It was indicated the quality of assembly and annotations of the chicken genome need to be improved and more researches are required for improving our understanding on function of genes and metabolic pathways of avian species. We conclude that the chicken PGDB is useful for studies on avian and chicken metabolism and provides a platform for comparative genomic and metabolic analysis of animal biology and biomedicine.

(Key words : metabolic reconstruction, chicken genome, pathway genome database, bioinformatics)

## 서론

닭(*Gallus gallus*)은 인류에게 중요한 단백질 공급원으로써 축산업에 있어 중요한 가축이며, 조류인 닭은 사람을 기준으로 생쥐 등의 실험 동물보다 먼저 종 분화가 이루어졌기 때문에 비교 생물학적 접근을 통해 의학, 생명의 진화 및 발생학적 측면 연구의 모델 동물로서 그 가치가 높다(Seo et al., 2009). 따라서 닭의 대사 생리에 대한 연구는 생산성, 질병 저항성, 사료 효율 등과 관련된 산업적 가치가 높을 뿐만 아니라 생물학, 의학적으로도 매우 중요하다.

동물의 대사는 효소에 의한 반응과 세포의 유기체 전달 및 유기 화합물의 전환을 통해 생명 현상을 조절하는 복잡

한 과정으로, 유전체에 담긴 정보에 의해 발현, 조절되기 때문에 유전체 수준 모델(genome-scale model)을 통한 분석이 필요하다(Schilling et al., 1999). 유전체 수준 모델 분석은 단순히 유전자의 구조, 기능을 밝히는 것이 아니라, 유전자형과 표현형의 상관관계를 분석, 해석, 예측함을 말한다. 유전체 수준 모델 분석을 위해 대사 경로 재확립이 선행되어야 하며, 이에 따라 박테리아, 단순 진핵생물 또는 고등 진핵생물을 포함한 다양한 생물체의 유전체 대사 경로 구축이 시도되었다(Seo and Lewin, 2009). 특히, Pathway Tools 소프트웨어는 전체 대사 경로를 구축하는데 유용하게 활용될 수 있는데, PathoLogic 알고리즘을 이용하여, 동물 유전체의 정보와 MetaCyc(Caspi et al., 2006)에 포함된 대사 경로를 비

\* To whom correspondence should be addressed : swseo@cnu.kr

교, 분석하여 종 특이적 PGDB(Pathway Genome Database)를 생물정보학적으로 생성할 수 있도록 한다(Karp et al., 2010). 현재까지 다양한 박테리아, 식물 및 동물 유전체 정보를 활용한 673개의 종 특이적 PGDB가 구축되어 BioCyc collection (<http://biocyc.org>)에 '-Cyc' 데이터베이스의 형태로 포함되어 있다(Karp et al., 2005). 통일된 데이터베이스 구조를 갖는 PGDB는 중간 비교 분석이 용이하며, 이를 이용해 대사 경로 중심(metabolism-centered)의 유전체 주석(genome annotation)이 가능함이 보고되었다(Seo and Lewin, 2009). Pathway Tools 소프트웨어를 이용하여 경제 동물 중 소 특이 PGDB인 CattleCyc이 구축된 바 있다(<http://biocyc.org/CATTLE/>).

닭의 유전체 염기 서열 분석 결과는 2004년에 처음 발표되었고, 2006년 5월에 발표된 Gallus-gallus build 2.1이 현재까지 사용되고 있다. 유전체의 크기가 포유동물에 비해 30% 정도인 1.2 Gb 가량이고, 유전자 수는 약 2만개 이상인 것으로 추정되고 있다(Hillier et al., 2004). 닭의 유전체 정보에 기반한 대사 경로의 재구성은 Reactome Project를 통해 이루어지고 있으나(<http://www.reactome.org>), 아직까지 '-Cyc' 데이터베이스 형태의 PGDB가 구축된 바 없다. 따라서 본 연구의 목적은 인터넷 상의 생물 정보 데이터베이스를 통해 제공되고 있는 닭 유전체 및 단백질, 대사체 정보를 바탕으로 대사 경로를 재확립하고, 닭 특이 PGDB를 구축함에 있다. 이를 위해 NCBI(National Center for Biotechnology Information; Maglott et al., 2007)에서 제공하는 유전자 모델 및 유전체의 구조적, 기능적 정보와 UniProt(Universal Protein Resource; Apweiler et al., 2010)의 단백질 정보 및 KEGG(Kyoto Encyclopedia of Genes and Genomes; Kanehisa et al., 2006)의 대사 경로 관련 정보를 포함한 통합 데이터베이스를 구축하고, Pathway Tools 소프트웨어를 이용해 유전체 주석과 MetaCyc의 대사 경로 정보를 연결하여 생물정보학적으로 닭 특이 대사 경로 재확립 및 PGDB를 구축하였다.

## 방 법

### 1. 닭 특이 통합 데이터베이스 구축

닭의 유전체 정보에 기반한 대사 경로 재확립은 유전자 특히 단백질 발현 유전자의 기능에 관한 정보를 필요로 한다. 보다 완전한 genome annotation(유전체 주석)을 위해 Perl 언어를 기반으로 개발된 자동 파이프라인(pipeline)(Kim et al., 2010)을 이용하여 여러 생물정보 데이터베이스에 산재해

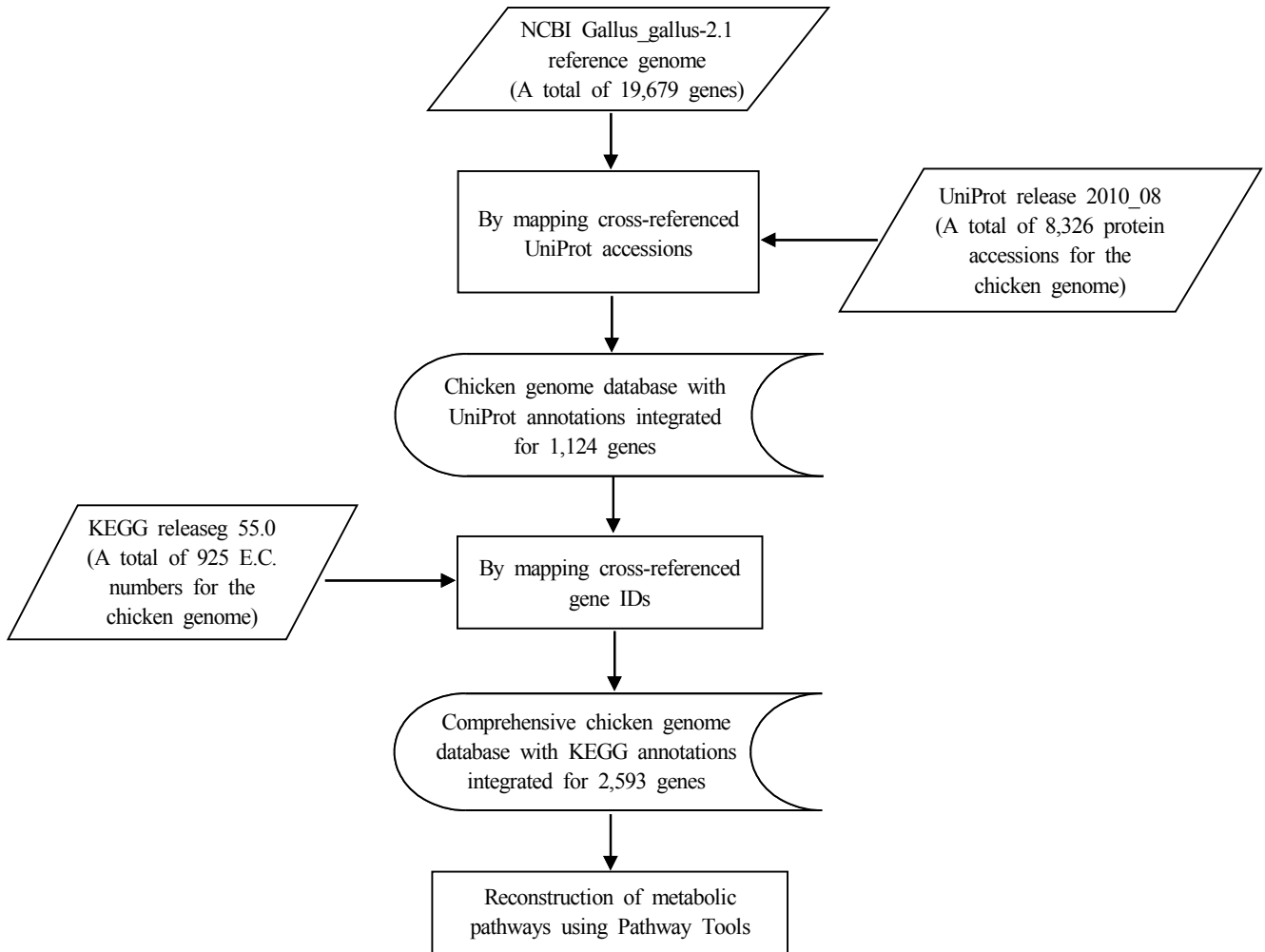
있는 정보를 통합한 닭 특이 통합 데이터베이스를 구축하였다. 파이프라인에 의한 유전체 정보 통합과정을 요약하면 다음과 같다. 1) NCBI와 Ensembl에 제공되는 종별 유전체 정보(유전자 모델 및 주석)를 각각 FTP(<ftp://ncbi.nlm.nih.gov/gene/DATA>)와 BioMart(<http://www.biomart.org>; Durinck et al., 2005)를 통해 자동 다운로드, 2) UniProt의 단백질 주석을 BioMart를 통해 다운로드 한 후, NCBI와 Ensembl genome annotation에 포함된 단백질 accession에 매핑(mapping)시켜 UniProt의 정보 통합, 3) KEGG의 대사 관련 정보를 FTP를 통해 다운로드 한 후, UniProt 정보가 통합된 NCBI 기반 genome annotation과 Entrez Gene ID를 매핑하여 통합, 4) 각각 NCBI 기반, Ensembl 기반 통합 genome annotation을 매칭(matching)과정을 통해 하나의 통합 데이터베이스(amalgamated genome database)로 구축한다. 다만, 본 연구에서는 NCBI와 Ensembl의 유전자 모델 및 scaffold 구성에 큰 차이를 보여 4번째 단계인 매칭을 통한 통합 데이터베이스는 구축하지 못하고, NCBI 유전자 모델에 기반한 닭 특이 통합 데이터베이스를 구축하였다(Fig. 1). 닭 유전체의 염기서열 및 기본 주석은 2010년 7월 19일자 NCBI Gallus\_gallus-2.1 reference genome을 기반으로 하고 단백질의 정보는 2010년 7월 13일자 UniProt release 2010\_08, 효소 및 대사관련 KEGG의 정보는 2010년 7월 1일자 release 55.0을 이용하였다.

위의 과정을 통해 구축된 통합 데이터베이스는 각 유전자별 염색체상의 위치, strand, biotype(protein coding, pseudo, microRNA 등), 유전자 이름, 단백질 이름, 기능 설명, EC(Enzyme Commission) number, 여러 데이터베이스의 유전자 및 단백질 ID, Unigene ID, GO(Gene Ontology) ID, Pubmed ID 등의 정보를 가지고 있다.

### 2. 닭 특이 PGDB(Pathway Genome Database) 구축 및 대사 경로 재확립

구축된 닭 특이 통합 데이터베이스를 이용하여 닭 특이 대사 경로·유전체 데이터베이스(PGDB)를 구축하고, PathoLogic 알고리즘을 구현한 Pathway Tools 소프트웨어 version 13.5(Karp et al., 2010)을 이용하여 닭 특이 대사 경로를 재확립하였다. PathoLogic 알고리즘은 유전자의 기능에 관한 주석이 포함된 데이터베이스에 담긴 각 유전자의 EC number, 유전자가 발현하는 단백질의 기능 설명 및 GO ID 정보와 MetaCyc 데이터베이스 내에 존재하는 curate된 대사 경로를 비교 분석하여 종 특이 대사 경로를 재확립할 수 있도록 한다(Karp et al., 2010).

재확립된 닭 특이 대사 경로를 '-Cyc' 기반으로 작성된 사



**Fig. 1.** Scheme of developing a comprehensive chicken genome database using the pipeline and metabolic reconstruction for the chicken genome.

람, 생쥐, 소 특이 대사 경로와 비교하였다. 사람과 생쥐는 기존에 개발된 HumanCyc(Romero et al., 2004) version 13.5와 MouseCyc(Evsikov et al., 2009) version 1.36을 이용하였고, 소의 경우에는 manual curation이 더해지지 않은 상태에서의 두 종간 비교를 위해 Btau\_4.0을 기반으로 한, 소 특이 대사 경로를 새로이 구축하여 닭 특이 대사 경로와 비교 분석하였다.

## 결과 및 고찰

### 1. 닭 특이 통합 데이터베이스 구축

NCBI Gallus\_gallus-2.1의 유전자 모델 및 scaffold 구성을 기본 모델로 한 닭 특이 통합 데이터베이스는 미토콘드리아, 1~28번의 상동 염색체, Z와 W의 성염색체 및 3,008개의 con-

tig로 구성되어 있다. 총 19,679개의 유전자를 포함하고 있으며, 이 중 18,457개의 유전자는 단백질을 발현하는 것으로 예측되고, 그 밖에 1,044개의 pseudo gene, 154개의 miscellaneous RNA, 22개의 unknown, 2개의 other 타입으로 구성되어 있다.

유전자들 중 Refseq 또는 UniProt accession을 보유하고 있는 유전자는 18,201개이며, Unigene ID를 가지고 있는 유전자는 13,104개이다. 또한, 2,464개와 3,277개의 유전자가 각각 EC number, KEGG pathway에 대한 정보를 보유하고 있고, GO term의 경우 biological process, cellular component, molecular function에 관한 정보를 각각 6,262, 5,882, 8,118개의 유전자가 가지고 있다.

특히, 유전자의 이름 및 기능 설명은 NCBI과 UniProt의 주석을 모두 포함하고 있기 때문에, 동일한 유전자가 여러 가

지로 명명됨에 따른 혼란을 최소화 하였다. 본 연구의 결과로 구축된 닭 특이 통합 데이터베이스의 정보는 웹 서비스의 형태로 공개될 예정이며, 현재 웹사이트가 구축 중에 있다.

기존에 구축된 소 유전체 통합 데이터베이스(Seo and Lewin, 2009)와 마찬가지로 닭 유전체 통합 데이터베이스도 NCBI 뿐만 아니라 Ensembl의 genome annotation도 통합하고자 하였다. 하지만, 닭 유전체 데이터베이스의 경우는 소 유전체와는 달리 NCBI와 Ensembl 데이터베이스가 제공하는 유전자 모델과 scaffold 구성의 차이가 큰 관계로 두 데이터베이스를 통합할 수 없었다. NCBI에서 제공하는 닭 유전체의 염색체 및 un-assigned scaffold의 수는 앞에서 언급한 바와 같이 총 3,039개인데 반해, Ensembl 데이터베이스에서 제공하는 염색체 및 scaffold는 미토콘드리아, 1~28번의 상동 염색체, 1, 2, 4, 5, 7, 8, 10, 13, 16, 18, 20, 22, 25, 28번 염색체 각각의 random assembly, 2개의 연관 그룹과 그들의 random assembly, Z와 W의 두 개의 성염색체와 그들의 random assembly, unassigned contig의 random assembly로 총 52개였다. 기본 유전자 모델 및 주석을 NCBI를 기반으로 한 이유는 다른 데이터베이스(특히 대사 경로 및 효소 반응에 대한 정보를 담고 있는 KEGG 데이터베이스)의 내용을 통합하는데 NCBI 유전자 모델이 더욱 용이하고, 또한 NCBI의 경우 고유한 유전자 ID를 부여하는 반면 Ensembl의 경우 유전자 ID가 일정하지 않기 때문에 향후 데이터베이스를 업데이트함에 있어 NCBI 유전자 모델이 더 용이하기 때문이다.

본 연구에서 구축된 닭 특이 통합 데이터베이스는 Ensembl의 genome annotation을 통합하지는 못했지만, 10,975개의 유전자는 NCBI genome annotation에서 제공된 Ensembl gene ID를 보유하고 있으며, 유전자 이름 및 기능 설명 등의 내용은 UniProt 데이터베이스의 통합 과정에서 일부 병합이 되었으므로 닭 유전체에 관한 다양하고 많은 정보를 포함하고 있다고 할 수 있다. 향후 NCBI와 Ensembl간 닭 유전체에 대한 genome assembly 및 annotation의 통일성을 보다 더 높이는 노력이 필요한 것으로 생각된다.

## 2. 닭 특이 PGDB 구축 및 대사 경로 재확립

본 연구에서 Pathway Tools 소프트웨어를 이용하여 닭의 유전체 정보를 대사 경로와 연계한 대사 경로·유전체 데이터베이스를 구축하였는데, Pathway Tools 소프트웨어를 사용한 이유는 Seo and Lewin(2009)이 언급한 바와 같이 1) PathoLogic 알고리즘을 구현한 Pathway Tools 소프트웨어는 가장 포괄적이고 curation이 잘된 데이터베이스 중 하나인 MetaCyc 데이터베이스와 genome annotation을 매핑하여 보다 쉽

고 정확하게 유전자와 대사 경로가 연계된 종 특이적 대사 경로의 재확립을 가능하게 하며, 2) BioCyc에 소장되어 있는 673종의 PGDB들과 같이 '-Cyc' 형식의 데이터베이스 형식을 이용함으로써 동일한 플랫폼에서 다른 종과의 대사 경로 비교 분석이 용이하고(Urbanczyk-Wochniak and Sumner, 2007), 3) 재구축된 대사 네트워크를 시스템 생물학적 모델에 적용할 때 MetaCyc의 대사 경로 구현 방식이 KEGG에 비해 오류가 적고(Poolman et al., 2006), 4) Pathway Tools에 내재된 'Omics Viewer'를 통해 향후 유전자의 전사체, 단백질체, 대사체 그리고 reaction flux에 대한 데이터를 PGDB에 통합시키기 용이(Choi et al., 2007)하기 때문이다.

Pathway Tools를 이용하여 닭의 유전체 정보를 MetaCyc 데이터베이스에 포함된 1,530여 개의 대사 경로와 분석한 결과, 현재의 닭 유전체 Gallus gallus-2.1에서 2,709개의 효소, 71개의 운반체(transporter)와 1,698개의 효소반응, 8개의 운반 반응(transport reaction)이 도출되었다. 이를 통해 총 212개의 대사 경로가 재확립되었고, 1,360개의 화합물(compound)이 닭 특이 대사 데이터베이스에 포함되었다(Table 1). 총 대사 경로의 수는 사람(285), 생쥐(224), 소(298)에 비해 적었다.

닭 특이 212개의 대사 경로에 총 215개의 Pathway hole이 존재하는데(Table 1), Pathway hole이란 대사 경로 중 생화학적 반응을 매개하는 효소를 발현하는 유전자가 현재의 유전체에서 확인되지 않은 어느 특정 반응(reaction)을 말한다. 본 연구에서 구축된 닭 특이 대사 데이터베이스의 경우 총 215개의 반응이 그 반응을 매개하는 단백질 및 유전자의 존재가 아직 확인되지 않았다. Pathway hole이 존재하는 이유는 크게 두 가지로, 그 종이 특정 대사 경로를 가지고 있지 않거나, 유전체의 genome annotation에 문제가 있는 경우(유전체 assembly 오류 또는 annotation의 미비 또는 오류)이다. 대사 경로 재확립을 통한 pathway hole의 규명은 유전체의 assembly와 annotation의 질을 대변하기도 한다(Seo and Lewin, 2009).

닭 특이 대사 경로에 존재하는 215개의 pathway hole은 대사 경로의 전체 반응의 30%로 충분한 manual curation을 거친 대장균(6%, <http://ecocyc.org>), 소(15%, <http://biocyc.org/CATTLE/>) 보다는 높았으며, manual curation 정도가 대장균과 소보다는 낮은 생쥐(27%)와 비슷한 수준이었다(Table 1). 사람의 경우, 370개(44%)의 많은 pathway hole이 존재하는 데, 이는 사람 PGDB의 기본 방향이 더 많은 대사 경로를 도출하는 데 있음(personal communication with Dr. Peter D. Karp)에 기인한 것으로 생각된다. 도출된 대사 경로, 효소 반응, 효소의 개수 및 pathway hole의 비율(26%) 등의 측면에서, 본 연구에서 함께 재확립된 소 특이 PGDB와 비교할 때, 닭의 유전체 anno-

**Table 1.** Comparison of selected organism-specific pathway genome databases (PGDB)

Database statistics	Chicken	Human <sup>†</sup>	Mouse <sup>‡</sup>	Cattle
Metabolic pathways	212	285	224	298
Enzymatic reactions	1,698	1,598	1,686	2,040
Enzymes	2,709	2,821	11,101	4,116
Compounds	1,360	1,115	1,140	1,529
Pathway holes				
Number of Pathway holes	215	370	192	260
Percentage*	30%	44%	27%	26%
Pathway with No Holes	121	116	134	179
Pathways with 1 Hole	47	64	40	53
Pathways with 2 Holes	23	49	22	28
Pathways with 3 Holes	7	20	10	9
Pathways with 4 Holes	3	11	7	10
Pathways with 5 Holes	2	8	3	5
Pathways with >5 Holes	9	17	8	14
Total Pathway with Holes	91	169	90	119

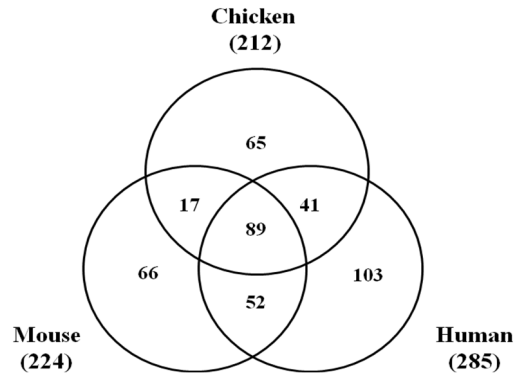
\*Pathway holes as percentage of total reactions in pathways.

<sup>†</sup> HumanCyc 13.5, <sup>‡</sup> MouseCyc 1.36.

tation의 완성도가 소에 비해 떨어짐을 알 수 있다(Table 1). 향후 닭 유전체 assembly 및 annotation의 질을 높이는 과정과 manual curation을 통하여 구축된 닭 특이 PGDB의 pathway hole을 줄이는 노력이 필요한 것으로 사료된다.

### 3. 대사 경로의 종 간 비교 분석

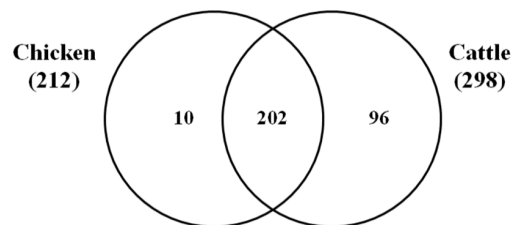
구축된 닭의 대사 경로를 기 발표된 사람과 생쥐의 대사 경로 및 새로 재확립한 소의 대사 경로와 비교 분석하였다. 사람, 생쥐, 닭의 대사 경로와 비교할 때, 닭에만 존재하는 특이 대사 경로는 65개이며, 사람과 생쥐는 각각 103개, 66개였다(Fig. 2). 닭과 사람 두 종간 동일 대사 경로는 130개이며, 닭과 생쥐의 경우는 106개의 대사 경로가 같았고, 닭, 사람, 생쥐의 공통 대사 경로는 89개로 나타났다(Fig. 2). 이러한 종 간의 차이는 실제 대사 경로의 차이이기 보다는 각 데이터 베이스가 구축되는 과정에서의 차이에서 기인한 것으로 보인다. Seo and Lewin(2009)의 소와 사람의 대사 경로 비교에서 나타난 바와 같이 데이터베이스가 구축되는 과정에서 manual curation 정도의 차이와 연구자간 일관적이지 않은



**Fig. 2.** Comparison of constructed metabolic pathways among the chicken, mouse and human genomes. Numbers of metabolic pathways, shared by chicken, HumanCyc 13.5 and MouseCyc 1.36. The numbers in parentheses are the total number of pathways in the corresponding databases.

대사 경로에 대한 정의로 인해 실제 종 간의 차이에서 기인하지 않은 데이터베이스 간의 차이가 존재한다. 향후 대사 경로 데이터베이스 간의 통일성을 높이는 노력이 필요한 것으로 생각된다.

닭에 특이적으로 존재하는 대사 경로를 규명하기 위해 재확립된 소의 대사 경로와 좀 더 심도 있는 비교 분석을 실시한 결과, 소의 대사 경로에는 존재하지 않는 닭 특이적인 10개의 대사 경로를 확인하였다(Fig. 3). 하지만, 이 중 6개의 대사 경로는 소의 대사 경로 구축 과정에서 유전자 주석의 미비로 데이터베이스에 포함되지 않은 대사 경로였으며, 4개의 경우, 닭의 대사 경로 구축 과정 중에 실제 닭에 존재하지 않음에도 false-positive로 포함된 대사 경로임이 밝혀졌다(Table 2). 결과적으로 닭의 대사 경로는 모두 소의 대사 경로에 포함되므로 두 종 간 대사 경로의 차이를 확인할 수 없었다. 이러한 문제점은 Seo and Lewin(2009)이 지적한 바와



**Fig. 3.** Comparison of constructed metabolic pathways between the chicken and cattle genomes. Numbers of metabolic pathways, shared by chicken and cattle. The numbers in parentheses are the total number of pathways in the corresponding databases.

같이 Metacyc에 포함되어 있는 대사 경로가 박테리아와 식물에 근간을 두기 때문인데, 이후 사람, 생쥐 등의 포유동물 대사 경로가 많이 첨가되긴 하였지만, 조류의 대사 경로에 대한 정보는 아직 미흡하기 때문인 것으로 보여진다. 따라서 본 연구에서 생물정보학적 기법을 이용해 재확립된 닭의 대사 경로를 바탕으로 조류 특히 닭에서 행해진 실험 결과를 토대로 한 manual curation은 앞으로의 연구에 필요한 과정이며, 이를 통해 닭 또는 조류 특이 대사 경로를 연구하는데 도움이 될 것으로 생각된다. 또한, 본 연구에서 개발된 닭 특이 PGDB 및 재확립된 대사 경로는 동일한 '-Cyc' 형식의 데이터베이스로 포유동물 및 미생물과의 비교 생물학적 접근을 통한 의학, 생명의 진화 및 발생학적 측면 연구에 활용될 것으로 기대된다.

## 적 요

닭의 대사 생리에 대한 연구는 산업적 가치 및 생물학, 의학적으로도 매우 중요하다. 닭의 유전체 염기서열 분석 결과는 2004년에 처음 발표되었고, 이러한 유전체 정보를 바탕으로 유전형과 표현형의 상관관계를 분석하는 연구가 필요하

다. 따라서 본 연구는 닭 유전체 정보를 바탕으로 대사 경로를 재확립하고, 닭 특이 대사 경로·유전체 데이터베이스를 구축하였다. 이를 위해 Perl 언어를 기반으로 개발된 자동 파이프라인(pipeline)을 이용하여 여러 생물정보 데이터베이스에 산재해 있는 닭 유전체에 관한 정보를 통합한 닭 특이 통합 데이터베이스를 구축하였다. 또한, 구축된 닭 특이 통합 데이터베이스를 기반으로 PathoLogic 알고리즘을 구현한 Pathway Tools 소프트웨어를 이용하여 닭 특이 대사 경로를 재확립하였다. 결과적으로, 닭 유전체 Gallus\_gallus-2.1에서 2,709개의 효소, 71개의 운반체(transporter)와 1,698개의 효소 반응, 8개의 운반 반응(transport reaction)이 도출되었다. 이를 통해 총 212개의 대사 경로가 재확립되었고, 1,360개의 화합물(compound)이 닭 특이 대사 데이터베이스에 포함되었다. 다른 종(사람, 생쥐, 소)과의 비교 분석을 통해 중요한 대사 경로가 닭 유전체에 보존되어 있음을 보였다. 또한, 닭 유전체의 assembly와 annotation의 질을 높이는 노력과 닭 및 조류에서 유전자 기능 및 대사 경로에 대한 연구가 필요한 것으로 나타났다. 결론적으로, 본 연구에서 재확립된 닭의 대사 경로 및 데이터베이스는 닭 및 조류의 대사 연구뿐만 아니라 포유동물 및 미생물과의 비교 생물학적 접근을 통한 의학 및 생물학적 연구에 활용될 것으로 기대된다.

**Table 2.** List of the ten chicken pathways not found in the cattle metabolic reconstruction

Pathway	Status	Comment
$\beta$ -alanine biosynthesis V	Missing in Cattle	The enzymatic reaction, 1.5.3.- was not annotated for SMOX in the cattle database.
2-amino-3-carboxymuconate semialdehyde degradation to glutaryl-CoA	Missing in Cattle	Omitted in the cattle metabolic reconstruction because it is redundant with 'tryptophan degradation III (eukaryotic) mammals'.
Dolichol and dolichyl phosphate biosynthesis	Missing in Cattle	Dehydrololichyl diphosphate synthase reaction was not annotated in the cattle genome. Nonetheless, there is no strong evidence of presence or absence of this pathway in chicken.
Alanine degradation III	Missing in Cattle	It is annotated as 'alanine biosynthesis' pathway in the cattle database (redundant pathway).
Putrescine biosynthesis III	Missing in Cattle	Omitted in the cattle metabolic reconstruction because it is redundant with 'putrescine biosynthesis IV'.
Tryptophan degradation to 2-amino-3-carboxymuconate semialdehyde	Missing in Cattle	Omitted in the cattle metabolic reconstruction because it is redundant with 'tryptophan degradation III (eukaryotic) mammals'.
$\beta$ -carboline biosynthesis	False-positive	It is a plant pathway and not present in human, mouse, or cattle.
Cardiolipin biosynthesis II	False-positive	It is a bacterial pathway and not present in human, mouse, or cattle.
Methionine biosynthesis IV	False-positive	Methionine is not synthesized in chicken.
Oleate $\beta$ -oxidation	False-positive	It is a bacterial pathway and not present in human, mouse, or cattle.

(색인어 : 대사 경로 재확립, 닭 유전체, 대사 경로 · 유전체 데이터베이스, 생물정보학)

## 사 사

이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초 연구 사업(No. 2009-0064205)입니다.

## 인용문헌

- Apweiler R, Martin MJ, O'Donovan C, Magrane M, Alam-Faruque Y, Antunes R, Barrell D, Bely B, Bingley M, Binns D, Bower L, Browne P, Chan WM, Dimmer E, Eberhardt R, Fedotov A, Foulger R, Garavelli J, Huntley R, Jacobsen J, Kleen M, Laiho K, Leinonen R, Legge D, Lin Q, Liu WD, Luo J, Orchard S, Patient S, Poggioli D, Pruess M, Corbett M, di Martino G, Donnelly M, van Rensburg P, Bairoch A, Bougueleret L, Xenarios I, Altairac S, Auchincloss A, Argoud-Puy G, Axelsen K, Baratin D, Blatter MC, Boeckmann B, Bolleman J, Bollondi L, Boutet E, Quintaje SB, Breuza L, Bridge A, deCastro E, Ciapina L, Coral D, Coudert E, Cusin I, Delbard G, Doche M, Dornevil D, Roggli PD, Duvaud S, Estreicher A, Famiglietti L, Feuer-mann M, Gehant S, Farriol-Mathis N, Ferro S, Gasteiger E, Gateau A, Gerritsen V, Gos A, Gruaz-Gumowski N, Hinz U, Hulo C, Hulo N, James J, Jimenez S, Jungo F, Kappler T, Keller G, Lachaize C, Lane-Guermonprez L, Langendijk-Genevaux P, Lara V, Lemerrier P, Lieberherr D, Lima TD, Mangold V, Martin X, Masson P, Moinat M, Morgat A, Mottaz A, Paesano S, Pedruzzi I, Pilbout S, Pillet V, Poux S, Pozzato M, Redaschi N, Rivoire C, Roechert B, Schneider M, Sigrist C, Sonesson K, Staehli S, Stanley E, Stutz A, Sundaram S, Tognolli M, Verbregue L, Veuthey AL, Yip LN, Zuletta L, Wu C, Arighi C, Arminski L, Barker W, Chen CM, Chen YX, Hu ZZ, Huang HZ, Mazumder R, McGarvey P, Natale DA, Nchoutmboube J, Petrova N, Subramanian N, Suzek BE, Ugochukwu U, Vasudevan S, Vinayaka CR, Yeh LS, Zhang J 2010 The Universal Protein resource (UniProt) in 2010. *Nucleic Acids Res* 38:D142-D148.
- Caspi R, Foerster H, Fulcher CA, Hopkinson R, Ingraham J, Kaipa P, Krummenacker M, Paley S, Pick J, Rhee SY, Tissier C, Zhang P, Karp PD 2006 MetaCyc: A multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res* 34:D511-D516.
- Choi C, Munch R, Leupold S, Klein J, Siegel I, Thielen B, Benkert B, Kucklick M, Schobert M, Barthelmes J, Ebeling C, Haddad I, Scheer M, Grote A, Hiller K, Bunk B, Schreiber K, Retter I, Schomburg D, Jahn D 2007 SYSTOMONAS - An integrated database for systems biology analysis of Pseudomonas. *Nucleic Acids Res* 35:D533-D537.
- Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazzma A, Huber W 2005 BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics* 21:3439-3440.
- Evsikov AV, Dolan ME, Genrich MP, Patek E, Bult CJ 2009 MouseCyc: A curated biochemical pathways database for the laboratory mouse. *Genome Biol* 10:R84.
- Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, Bork P, Burt DW, Groenen MAM, Delany ME, Dodgson JB, Chinwalla AT, Cliften PF, Clifton SW, Delehaunty KD, Fronick C, Fulton RS, Graves TA, Kremitzki C, Layman D, Magrini V, McPherson JD, Miner TL, Minx P, Nash WE, Nhan MN, Nelson JO, Oddy LG, Pohl CS, Randall-Maher J, Smith SM, Wallis JW, Yang SP, Romanov MN, Rondelli CM, Paton B, Smith J, Morrice D, Daniels L, Tempest HG, Robertson L, Masabanda JS, Griffin DK, Vignal A, Fillon V, Jacobsson L, Kerje S, Andersson L, Crooijmans RPM, Aerts J, van der Poel JJ, Ellegren H, Caldwell RB, Hubbard SJ, Grafham DV, Kierzek AM, McLaren SR, Overton IM, Arakawa H, Beattie KJ, Bezzubov Y, Boardman PE, Bonfield JK, Croning MDR, Davies RM, Francis MD, Humphray SJ, Scott CE, Taylor RG, Tickle C, Brown WRA, Rogers J, Buerstedde JM, Wilson SA, Stubbs L, Ovcharenko I, Gordon L, Lucas S, Miller MM, Inoko H, Shiina T, Kaufman J, Salomonsen J, Skjoedt K, Wong GKS, Wang J, Liu B, Wang J, Yu J, Yang HM, Nefedov M, Koriabine M, deJong PJ, Goodstadt L, Webber C, Dickens NJ, Letunic I, Suyama M, Torrents D, von Mering C, Zdobnov EM, Makova K, Nekrutenko A, Elnitski L, Esvara P, King DC, Yang S, Tyekucheva S, Radakrish-

- nan A, Harris RS, Chiaromonte F, Taylor J, He JB, Rijnkels M, Griffiths-Jones S, Ureta-Vidal A, Hoffman MM, Severin J, Searle SMJ, Law AS, Speed D, Waddington D, Cheng Z, Tuzun E, Eichler E, Bao ZR, Flicek P, Shteynberg DD, Brent MR, Bye JM, Huckle EJ, Chatterji S, Dewey C, Pachter L, Kouranov A, Mourelatos Z, Hatzigeorgiou AG, Paterson AH, Ivarie R, Brandstrom M, Axelsson E, Backstrom N, Berlin S, Webster MT, Pourquie O, Reymond A, Ucla C, Antonarakis SE, Long MY, Emerson JJ, Betran E, Dupanloup I, Kaessmann H, Hinrichs AS, Bejerano G, Furey TS, Harte RA, Raney B, Siepel A, Kent WJ, Haussler D, Eyraas E, Castelo R, Abril JF, Castellano S, Camara F, Parra G, Guigo R, Bourque G, Tesler G, Pevzner PA, Smit A, Fulton LA, Mardis ER, Wilson RK 2004 Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695-716.
- Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M 2006 From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* 34:D354-D357.
- Karp PD, Ouzounis CA, Moore-Kochlacs C, Goldovsky L, Kaipa P, Ahren D, Tsoka S, Darzentas N, Kunin V, Lopez-Bigas N 2005 Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res* 33: 6083-6089.
- Karp PD, Paley SM, Krummenacker M, Latendresse M, Dale JM, Lee TJ, Kaipa P, Gilham F, Spaulding A, Popescu L, Altman T, Paulsen I, Keseler IM, Caspi R 2010 Pathway Tools version 13.0: Integrated software for pathway/genome informatics and systems biology. *Brief Bioinform* 11:40-79.
- Kim WS, Lee SY, Seo S 2010 Development of an amalgamated cattle genome database based on Btau\_4.0. Page 30 In: Proceedings of the 14<sup>th</sup> AAAP Animal Science Congress, Asian-Australasian Association of Animal Production Societies, Pingtung, Taiwan, ROC.
- Maglott D, Ostell J, Pruitt KD, Tatusova T 2007 Entrez gene: gene-centered information at NCBI. *Nucleic Acids Res* 35: D26-D31.
- Poolman MG, Bonde BK, Gevorgyan A, Patel HH, Fell DA 2006 Challenges to be faced in the reconstruction of metabolic networks from public databases. *IEE Proc Syst Biol* 153:379-384.
- Romero P, Wagg J, Green ML, Kaiser D, Krummenacker M, Karp PD 2004 Computational prediction of human metabolic pathways from the complete human genome. *Genome Biol* 6:R2.
- Schilling CH, Schuster S, Palsson BO, Heinrich R 1999 Metabolic pathway analysis: Basic concepts and scientific applications in the post-genomic era. *Biotechnol Prog* 15:296-303.
- Seo S, Lewin HA 2009 Reconstruction of metabolic pathways for the cattle genome. *BMC Syst Biol* 3:33.
- Seo S, Paek WK, Lee JH 2009 Recent status of chicken genome researches. *Korean J Poult Sci* 36:111-115.
- Urbanczyk-Wochniak E, Sumner LW 2007 MedicCyc: a biochemical pathway database for *Medicago truncatula*. *Bioinformatics* 23:1418-1423.

(접수: 2010. 9. 7, 수정: 2010. 9. 18, 채택: 2010. 9. 20)