

Construction of Gene Network System Associated with Economic Traits in Cattle

Dajeong Lim¹, Hyung-Yong Kim², Yong-Min Cho¹, Han-Ha Chai¹, Jong-Eun Park¹, Kyu-Sang Lim¹ and Seung-Su Lee^{1*}

¹National Institute of Animal Science, RDA, Wanju 55365, Korea

²Insilicogen Inc., Suwon 441-813, Gyeonggi-do, Korea

Received June 19, 2016 / Revised July 26, 2016 / Accepted August 23, 2016

Complex traits are determined by the combined effects of many loci and are affected by gene networks or biological pathways. Systems biology approaches have an important role in the identification of candidate genes related to complex diseases or traits at the system level. The gene network analysis has been performed by diverse types of methods such as gene co-expression, gene regulatory relationships, protein-protein interaction (PPI) and genetic networks. Moreover, the network-based methods were described for predicting gene functions such as graph theoretic method, neighborhood counting based methods and weighted function. However, there are a limited number of researches in livestock. The present study systemically analyzed genes associated with 102 types of economic traits based on the Animal Trait Ontology (ATO) and identified their relationships based on the gene co-expression network and PPI network in cattle. Then, we constructed the two types of gene network databases and network visualization system (<http://www.nabc.go.kr/cg>). We used a gene co-expression network analysis from the bovine expression value of bovine genes to generate gene co-expression network. PPI network was constructed from Human protein reference database based on the orthologous relationship between human and cattle. Finally, candidate genes and their network relationships were identified in each trait. They were typologically centered with large degree and betweenness centrality (BC) value in the gene network. The ontle program was applied to generate the database and to visualize the gene network results. This information would serve as valuable resources for exploiting genomic functions that influence economically and agriculturally important traits in cattle.

Key words : Cattle, Gene co-expression network, potein-protein network, web database

서 론

유전자 네트워크는 유전자와 유전자가 기능적 혹은 물리적으로 연결되어 표현형이나 형질에 대하여 하나의 기능으로 설명될 수 있기 때문에 최근 생물학 연구에 매우 유용하게 사용된다. 특히, 유전자 네트워크 상에서 유전체를 기반으로 하여 하나의 유전자가 다양한 역할을 수행하기도 하며, 다수의 유전자들이 상호 연관관계를 맺어서 하나의 기능을 수행하기도 한다[2]. 예를 들면, 성격이 비슷한 사람들끼리 서로 가깝게 지내듯 기능이 유사한 유전자들도 밀접하게 관련되어 있어 이 관련성을 지도화한 유전자 네트워크를 활용하면 밀접하게 연관된 이미 밝혀진 이웃 유전자들의 기능을 통해 밝혀지지 않은 유전자의 기능을 예측할 수 있게 된다. 생물체 내에서

일어나는 모든 생리현상은 하나의 유전자가 아닌 유전자 여러 개의 상호작용에 의해 이뤄지기 때문이다. 유전자 네트워크는 유전자를 점으로 나타내며, 유전자와 유전자 사이는 연결점 또는 선으로 구성될 수 있다. 유전자 네트워크의 종류는 크게 발현, 단백질 상호작용, 전사조절 네트워크 등으로 구분할 수 있다. 발현 네트워크는 유전자 발현 데이터를 활용하여 특정 환경 또는 형질에서 발현 차이를 보이는 유전자들을 대량 발굴하여, 유전자들 사이에서 공발현을 가지고 있는 관계를 규명하게 된다[4]. 단백질상호작용 네트워크에서 각 점은 단백질 그리고 두 점 사이의 선은 단백질들 간의 상호작용 특히, 물리적 접촉을 의미하게 된다[22]. 전사조절 네트워크는 조절인자와 표적인자의 관계로 설명하는 네트워크이다[9]. 유일하게 네트워크상에 방향성을 가지고 있다. 예를 들면, 특정 대사경로에 참여하는 단백질군의 발현은 그들에게 공통적으로 특이성을 갖는 전사조절인자에 의해 결정되는 경우가 많다. 이들 전사조절인자들과 그들의 표적유전자들 간의 종속관계를 방향성을 가진 선들로 나타낸 것이 전사조절 네트워크이다.

이러한 유전자 네트워크 연구들이 가축의 경제형질과 관련하여 주요 대사회로 규명 및 유전자 네트워크 내의 허브 유전자 탐색을 통하여 분자유전학적 관점에서 진행되고 있다. 유전자 네트워크 및 결과에 대한 대용량 정보를 제공하는 웹

*Corresponding author

Tel : +82-41-580-3358, Fax : +82-41-580-3369

E-mail : genemap@korea.kr

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

시스템은 대표적으로 STRING 데이터베이스[23]이며, 단백질-단백질 상호작용 네트워크에 대하여 기능적, 물리적 상호작용 관계를 가중치 점수화한 결과를 보여준다. 뿐만 아니라, 마이크로어레이 자료를 활용하여 유전자와 유전자 사이의 공발현 정도를 사람, 쥐, 닭, 초파리, 꼬마선충 등에서 보여주는 웹 데이터베이스도 제공되고 있다[17]. 유전자 수준에서 유전자 네트워크 정보를 보여주는 웹 데이터베이스는 소수 존재하지만[13], 이를 형질과 연관시켜 분석한 결과를 제공하는 웹 데이터베이스는 현재까지 없는 실정이다.

가축 분야에서는 소에서 골격근 형성에 영향을 주는 후보 유전자를 탐색하기 위하여 47개의 전사체 자료를 이용한 유전자 발현 네트워크 분석을 통하여 myozenin-1 (MYOZ1), titin cap protein (TCAP), and PDZ, LIM domain 3 (PDLIM3) 유전자를 규명하였으며[20], 소의 19개 주요 경제형질과 연관된 유전 네트워크를 구축하기도 하였다[11]. 뿐만 아니라, 국내에서는 유전자 네트워크 분석법을 활용하여 한우의 근내지방(마블링) 형성에 영향을 주는 heat shock protein β 1 (HSPB1) 유전자와 전사조절인자와의 네트워크 분석을 통하여 mitogen-activated protein kinase (MAPK) signaling pathway에 관여하는 유전자들이 HSPB1 유전자의 발현에 영향을 주는 것을 실험적 검증을 통해 최종 확인하였다[14].

본 논문에서는 소의 단일형질에 대하여 유전자 네트워크를 구축하고 분석한 결과를 확대하여 102개 경제형질에 대하여 유전자 공발현 네트워크 및 단백질 상호작용 네트워크 분석결과에 대하여 데이터베이스를 구축하였으며, 웹 시스템으로 구현하였다. 이를 통해 축산분야에서 다양한 경제형질에 연관된 유전자군 탐색을 위해 중요한 유전자군을 위주로 접근 가능하여 실험 검증에 적용할 수 있게 되었다.

재료 및 방법

경제형질 자료 및 유전자군 추출

가축에서 각 형질에 대한 정의는 동물 형질 온톨로지 정보에 따라 분류하였다 [10]. 'Exterior', 'Health', 'Meat traits', 'Milk traits', 'Production', 'Reproduction'의 6개 대분류에 따라 17개 중분류, 102개 소분류로 경제형질이 나뉜다. 또한, 소의 경제형질 관련 유전자군을 추출하기 위하여 Animal QTL database (<http://www.genome.iastate.edu/cgi-bin/QTLdb/BT/index>)에서 'QTL location by bp' 파일을 활용하여 각 형질에 대한 물리적 위치를 파악하여, 해당 형질의 물리적 위치에 존재하는 유전자군을 추출하였다.

유전자 공발현 네트워크(Gene Co-expression Network) 구축

유전자 발현 네트워크는 유전자의 발현 패턴이 같이 변하게 되는 유전자들을 연결한 네트워크이다. 유전자 발현 micro-

array 데이터 전처리를 위하여 NCBI의 GEO (Gene Expression Omnibus) 데이터로부터 소의 마이크로어레이 정보를 다운로드 받는다(<http://www.ncbi.nlm.nih.gov/geo/>). 소의 마이크로어레이 정보는 GEO series (GSE)15544, GSE15342, GSE 13725, GSE6918, GSE10695, GSE12327, GSE9256, GSE12688, GSE11495, GSE11312, GSE7360, GSE9344, GSE8442이다. 모든 마이크로어레이의 CEL 파일들을 대상으로 바탕값 보정 및 표준화 등 데이터 전처리를 수행한다. 이는 R 통계프로그램의 'affy' 패키지를 활용하여 RMA (Robust Multiarray Average) 기법을 활용하였다[6]. 또한, RMA를 통해 얻어진 발현값은 로그값으로 변환시켜 추후 발현 네트워크 분석에 사용하였다. 공발현 네트워크는 유전자 하나하나의 노드로 표현되며, 유전자의 발현패턴이 비슷하다면 선으로 연결하여 표현하게 된다. 유전자 발현패턴의 비슷한 정도는 각각의 유전자에 대한 발현값들에 대하여 피어슨 상관관계수(r value)값을 계산하여 인접행렬을 구성한다. 공발현 네트워크는 가중치가 없는 네트워크를 사용하였다. 유전자의 발현패턴이 일정값 이상이 되면 두 개의 유전자가 연결되게 되며($r = 1$), 일정값 미만이 되면 두 유전자는 연결되지 않는($r = 0$) 두 가지의 정보로 구성된다. 공발현 네트워크는 공발현되는 유전자 그룹을 찾고, 해당 그룹에 찾기자 하는 형질과 관련성을 가지는지, 해당되는 유전자 그룹의 기능(예: 대사회로, 유전자 온톨로지 등)을 규명하는 것이 중요하다. 공발현되는 유전자 네트워크 분석은 R 통계 프로그램의 WGCNA (Weighted Gene Co-expression Network Analysis) 패키지에서 'hard thresholding' 알고리즘을 활용하여 각 형질에 대하여 공발현되는 유전자 그룹을 추출하였다[19].

단백질-단백질 상호작용(Protein-Protein Interaction) 네트워크 구축

단백질상호작용 네트워크에서 각 점은 단백질 그리고 두 점 사이의 선은 단백질들 간의 상호작용을 나타낸다. 주로 단백질 상호작용 정보는 효모에서 가장 많이 존재하며, 동물 특히 가축을 대상으로 단백질 상호작용 정보에 대한 데이터베이스는 존재하지 않는다. 따라서 본 연구에서는 human protein-protein interaction (HPRD) 정보를 활용하여 각 유전자에 대한 단백질 상호작용 정보를 추출하여 단백질 상호작용 네트워크를 구축하였다. HPRD는 기보고된 논문을 바탕으로 전문가에 의해 단백질 상호작용을 분석한 데이터베이스이다 [18]. 본 시스템에서는 소의 유전자에 대한 단백질 상호작용을 분석해야 하므로, NCBI에서 제공하는 Homologene 데이터베이스(<http://www.ncbi.nlm.nih.gov/homologene>)를 활용하여 중간 보존적인 그룹에 따라 HPRD에 존재하는 Human protein에 대응하는 소의 유전자를 찾은 후에, 관련 단백질 상호작용 정보를 추출하였다.

네트워크 토폴로지(Network topology) 분석

추출된 유전자 리스트를 대상으로 네트워크 전체의 토폴로지(topology)를 관찰하기 위해서 파이썬에서 제공하는 networkx 모듈을 활용하여 노드의 도수, 근접 중심성, 매개 중심성 값을 계산하였다. 노드의 도수는 유전자가 다른 유전자들과 연결된 정도를 의미한다. 유전자의 도수가 높을수록 해당 유전자는 다른 유전자들과 상호작용이 높다고 판단하기 때문에 네트워크 상에서 중요한 허브 유전자가 될 수 있는 가능성이 높다. 또한, 유전자의 매개중심성은 유전자들을 연결해주는 다리 역할을 해주므로, 이 역시 중요한 유전자일 가능성이 높다. 실제로 노드의 도수와 매개중심성 값은 매우 높은 상관관계를 가지고 있다. $g_{i \rightarrow j}$ 는 노드 i부터 j까지 노드 k를 경유하는 거리를 의미한다[3].

$$b(k) = \sum_{i,j} b_{i \rightarrow j}(k) = \sum_{i,j} \frac{g_{i \rightarrow j}^k}{g_{i \rightarrow j}}$$

근접중심성값은 노드 u부터 n-1까지 모든 경우 가능한 거리 중, 가장 짧은 거리를 의미한다. 노드의 수가 네트워크(그래프)의 거리에 영향을 주기 때문에, 근접성은 n-1까지 가장 최소값의 거리를 가지는 것을 선택하게 된다.

$$C(u) = \frac{n-1}{\sum_{v=1}^{n-1} d(v,u)}$$

또한, 유전자 네트워크는 power-law function을 따르게 된다. 대부분의 생물학적 네트워크에서 도수 분포는 멱함수 분포($p(k) \sim k^{-\lambda}$)를 따르고 있기 때문에 척도 없는 네트워크라고도 한다[1]. 각 형질에 대하여 power-law function을 활용하여 도수 분포도를 구성하여 해당법칙에 맞는지 확인하였다.

결과 및 고찰

유전자 네트워크 분석 시스템 구축

구축된 유전자 공발현 네트워크, 단백질-단백질 상호작용 네트워크 정보는 관계형 데이터베이스로 구축하고, 웹 인터페이스화하기 위하여, Ontle 프로그램을 이용하였다. Ontle은 “웹기반의 온톨로지 편집 운영 시스템”[12]으로, 네트워크 데이터를 관계형 데이터베이스의 노드, 엣지, 엣지타입 테이블로 관리하며, 특정한 노드와 인접한 노드를 연결 수준에 따라 하나의 웹 화면에 가시화하여 출력하고, 노드와 엣지정보를 편집할 수 있는 인터페이스를 제공한다. 본 연구에서는 소의 경제형질 유전자 네트워크 분석 시스템 목적에 맞도록 Ontle의 데이터 모델을 확장하여, 통합하였다(Fig. 1). Ontle은 유전자 공발현 네트워크, 단백질-단백질 상호작용 네트워크의 2개 레코드를 포함한다. BovineQtlTrait, EdgeInfo 테이블은 Ontle의 노드, 엣지와 각각 일대일로 대응되어, 네트워크 연산과

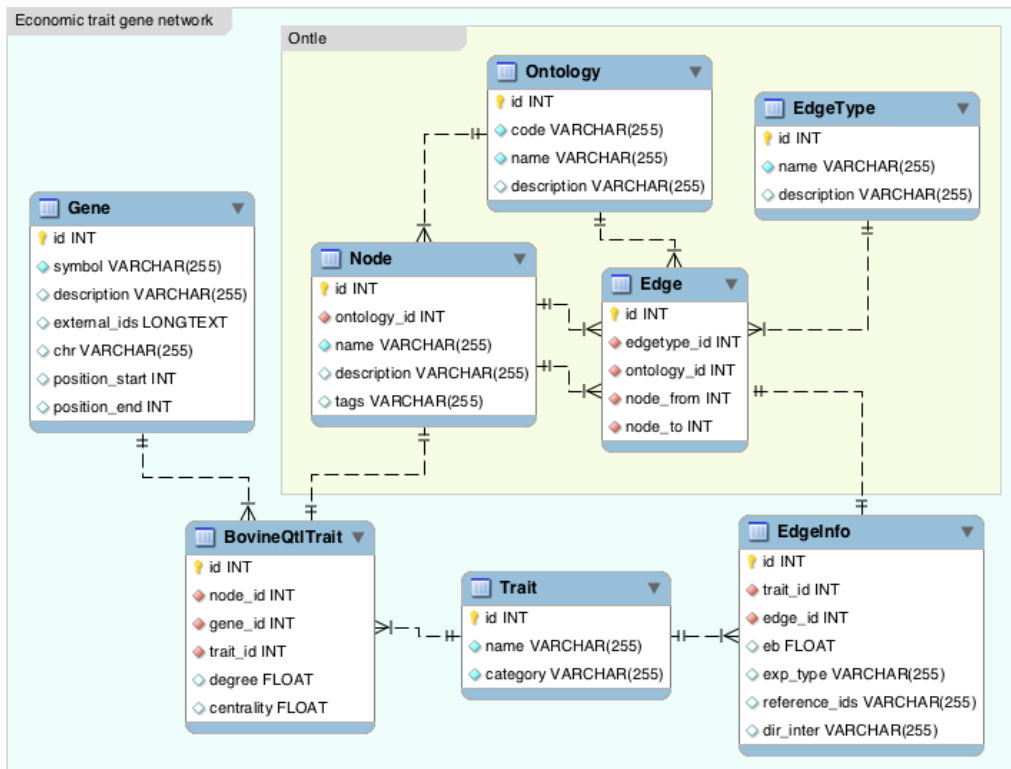


Fig. 1. Generation of tables from entities for gene network analysis system. “Ontle” is pre built solution for network data to RDB. “Economic trait gene network” extends Ontle’s RDB. Each box is DB table and the relations are displayed by line (1:N relation). Red key and red diamond is primary key and foreign key respectively.

관련된 기본 기능을 공유한다. 특정 유전자(Gene)는 양쪽 네트워크에 모두 포함될 수 있으며, 형질(Trait) 테이블을 참고하여, 특정 형질과 관련된 유전자를 탐색할 수 있다. 또한, 별도의 데이터로드 프로그램을 작성하여, 구축된 유전자 공발현 네트워크, 단백질-단백질 상호작용 네트워크 데이터를 OBO (Open Biomedical Ontologies) 형식으로 변환한 후[21], 본 모델에 맞도록 로드하고, 웹 사용자가 가시화된 네트워크 정보

를 탐색할 수 있도록 웹 어플리케이션을 구현하였다. 이를 위하여 파이썬 프로그래밍 언어와 장고(Django) 웹프레임워크, 마이스쿼(MySQL) 데이터베이스가 사용되었다. 사용자는 메인 메뉴를 통해 네트워크를 선택하고, 관심 형질을 선택한 후, 해당 형질과 관련된 유전자의 목록을 열람한다. 또한 직접 입력 폼을 통하여, 네트워크 데이터의 노드와 엣지를 직접 생성할 수 있다(Fig. 2). 특정 유전자를 선택하면, 해당 유전자와

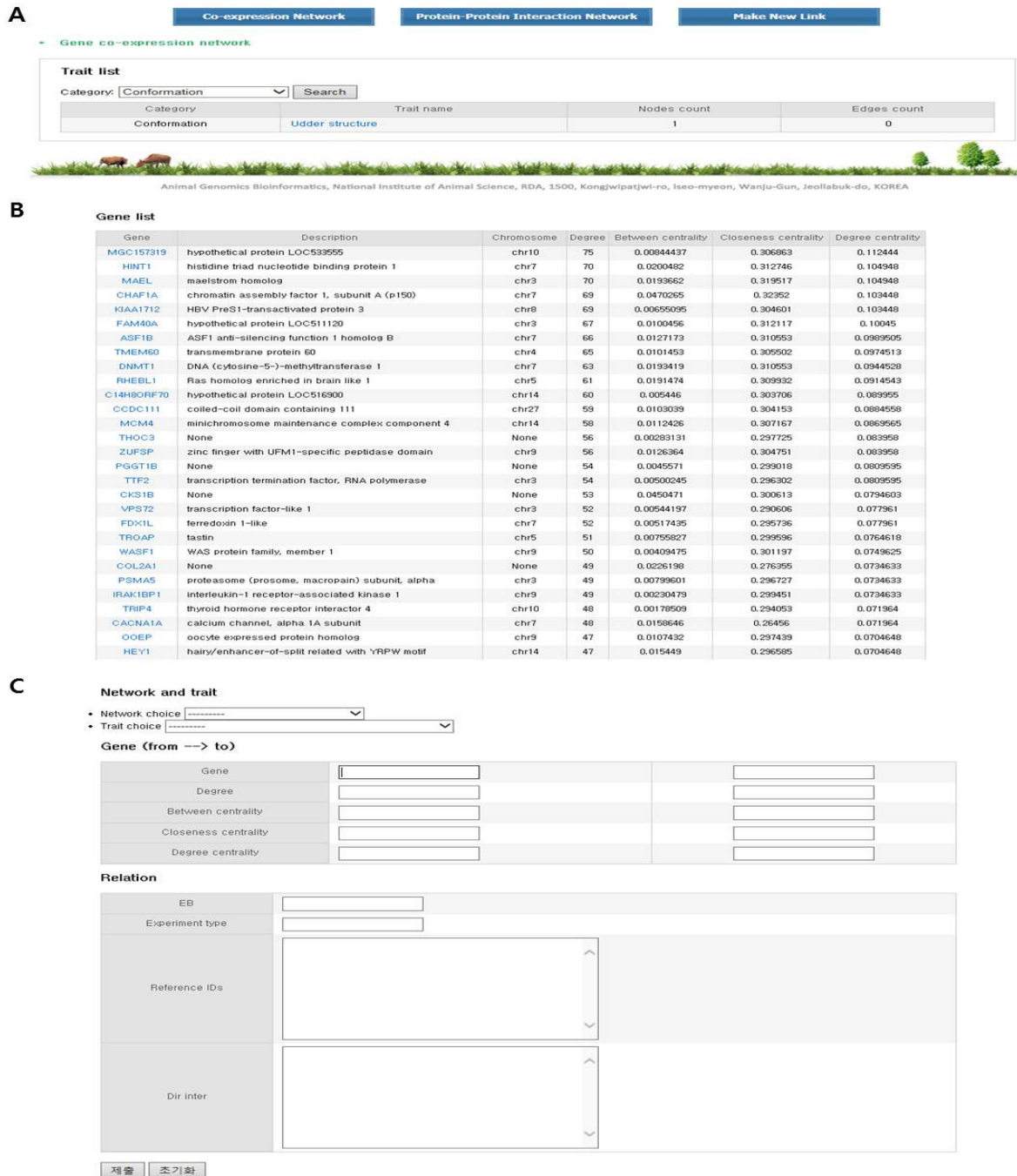


Fig. 2. The overview of bovine gene network analysis system. (A) The main page of the bovine gene network system (B) Genes list and network topology values for the specific trait (C) The input page of new gene network information.

네트워크를 통해 연결된 전체 유전자를 네트워크 형태로 가시화하여 출력한다. Fig. 3은 본 구축 시스템에서 네트워크 정보를 웹 가시화하는 방법과 화면 예제를 표시한다. 특정 노드(유전자) 선택시, 해당 노드와 연결된 노드를 그래프 연산을 통해 관계형 데이터베이스에서 추출한다. 사용자는 연결 수준을 지정할 수 있으며, level 2는 두 번 건너서 연결된 노드를, 레벨 3은 세 번 건너서 연결된 노드를 모두 추출한다. 추출된 네트워크는 DOT 그래프 형식[5]으로 변환된 후, 이미지 파일과 HTML 파일을 각각 생성하고 결합하여 화면에 출력한다. 출력 화면에서는 연결 가중치(강도)에 따라 엣지 색깔을 다르게 표시한다. 사용자는 연결 수준을 바꾸거나, 그래프 표시 방식을 바꾸거나, 노드/엣지 정보를 수정할 수 있다.

유전자 네트워크 데이터베이스 활용 분석 예시

본 시스템을 활용하여 한우의 마블링에 관련된 1) 양적형질 좌위 내에 있는 유전자들의 발현 패턴을 비교하여 2) 유전자의 공발현 형태를 관찰하고, 3) 가장 많이 연결되어 있는 허브 유전자군을 선별하여 실험 검증을 수행한 결과를 소개한다. 사용자는 소 유전체 네트워크 분석 시스템에서 ‘Co-expression Network’ 메뉴를 선택하여 Trait list의 category 선택창에서 ‘Meat Traits’의 ‘Marbling score’를 선택한다. 결과 화면은 ‘Marbling score’라는 유전좌위에 존재하는 유전자 리스트에 대하여 유전자 심볼, 유전자 설명, 염색체, 도수, 매개중심성, 근접중심성, 연결중심성 값이 제공된다. 사용자는 도수를 기준으로 관심있는 유전자에 대한 우선순위를 정하여 유전자 심볼을 클릭하여 해당 유전자와 공발현되는 유전자들이

어느 정도 연결되었는지 선 색의 진함 여부로 연결 강도를 살펴볼 수 있다. 도수를 기준으로 큰 도수를 갖는 6개 유전자와 큰 매개중심성 값을 가지는 5개 유전자를 마블링이 높은 그룹과 낮은 그룹에 대하여 실험 검증을 수행한 결과, TMEM 60과 DPYD 유전자가 두 그룹 사이에서 유전자 발현에 유의적인 차이($p < 0.05$)를 보이는 것을 확인할 수 있었다[15]. 뿐만 아니라, 본 시스템을 활용하여 단백질 상호작용 네트워크 분석을 통해 지방대사에 의해 한우의 근내지방 형성을 촉진시키는 허브 유전자군(PPAR γ , C/EBP α , RUNX1) 및 억제 유전자(RXRA, CAMK2A)를 규명하였다[16].

결론

과거에 매우 다양하고 활발한 생물학 연구에도 불구하고 여전히 기능이 검증되지 않은 유전자들이 많이 존재하고 있다. 유전자들의 다기능성을 고려한다면 전체 유전자 기능에서 우리가 현재 알고 있는 부분은 매우 한정적이게 된다. 이러한 유전자들의 기능을 유전자 네트워크 분석에 따라 효과적으로 예측할 수 있다면 전통적으로 진행되었던 후보유전자를 선별하여 연구하는 방법에 비해 비교적 단시간 내에 유전자의 잠재적 기능을 발굴 할 수 있을 것이다. 또한, 효모나 실험적 검증이 비교적 용이한 생물은 유전자 네트워크를 이용해 유전자의 기능상실이 초래할 형질변환의 결과도 예측하는 것이 가능할 수 있다. 많은 생물체에서 형질 또는 표현형들은 단일 생물대사회로나 혹은 서로 밀접하게 연결된 다수의 생물대사회로들에 의해 조절되기 때문이다. 동일한 생물대사회로에 속

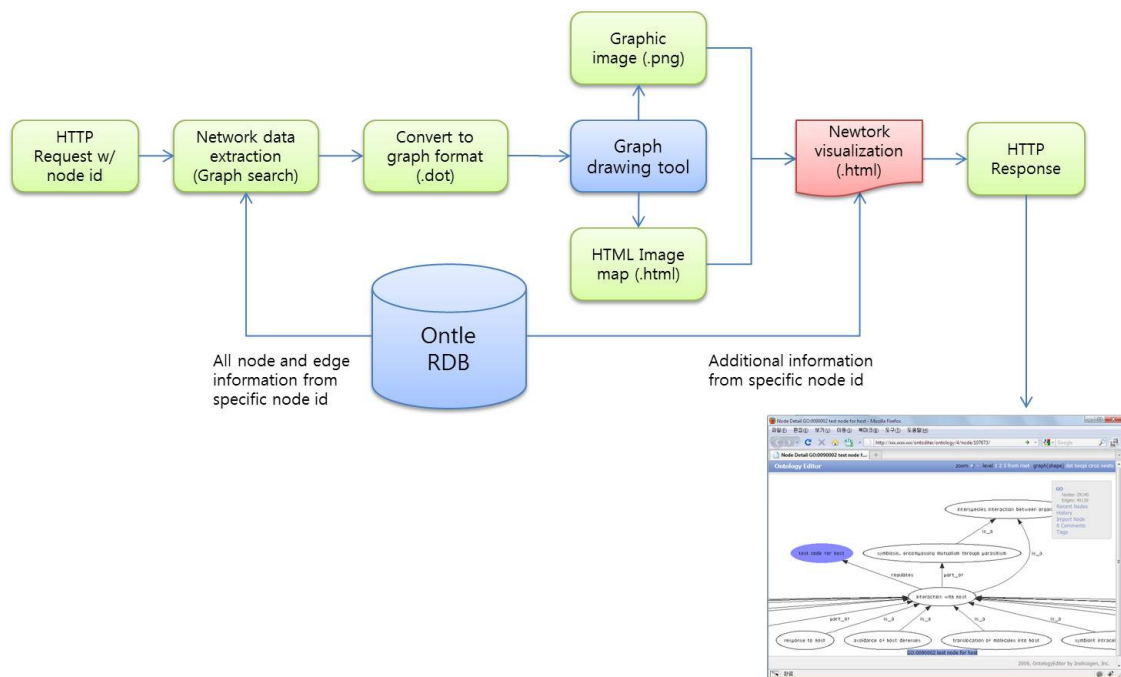


Fig. 3. The flowchart of gene network visualization based on the ontle program.

한 유전자들은 유전자 네트워크에서 서로 밀접하게 연결되어 비슷한 기능을 할 수 있다. 가축 역시 중요한 경제형질은 대부분 복합 형질로 구성되어 있으므로, 여러 대사회로나 조절경로에 관여하는 다수의 유전자 그룹이 동시에 발현되거나 기능을 나타내게 된다. 가축의 경제형질을 조절할 수 있는 원인이 되는 유전자들과 그들의 조절 유전자들을 순차적으로 밝히고 유전자들간의 복잡한 상관관계를 규명할 수 있게 된다. 본 시스템을 활용한 유전자 네트워크 접근법은 앞으로 축산분야에서 관심 있는 형질에 대한 후보 유전자의 수를 효과적으로 감소시켜줌으로써 신규 형질 후보유전자 선발에 강력한 도움을 줄 것으로 예상된다.

감사의 글

본 연구는 농촌진흥청 연구사업(PJ01098901)으로부터 연구비를 지원받아 수행하였으며, 연구비 지원에 감사드립니다.

References

- Barabasi, A. L. and Albert, R. 1999. Emergence of scaling in random networks. *Science* **286**, 509-512.
- Barabasi, A. L. and Oltvai, Z. N. 2004. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101-113.
- Brandes, U. 2001. A faster algorithm for betweenness centrality. *J. Math. Sociol.* **25**, 163-177.
- Carter, S. L., Brechbühler, C. M., Griffin, M. and Bond, A. T. 2004. Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics* **20**, 2242-2250.
- Ellson, J., Gansner, E., Koutsofios, L., North, S. C. and Woodhull, G. 2001. Graph Drawing . pp. 483-484. *Springer Berlin Heidelberg*, Germany.
- Gandhi ,T., Zhong, J., Mathivanan, S., Karthick, L., Chandrika, K., Mohan, S., Sharma, S., Pinkert, S., Nagaraju, S. and Periaswamy, B. 2006. Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. *Nat. Genet.* **38**, 285-293.
- Gautier, L., Cope, L., Bolstad, B. and Irizarry, R. 2004. af-fy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307.
- Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M., Hobbs, B. and Speed, T. P. 2003. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**, e15-e15.
- Hecker, M., Lambeck, S., Toepfer, S., Van Someren, E. and Guthke, R. 2009. Gene regulatory network inference: data integration in dynamic models – a review. *Biosystems* **96**, 86-103.
- Hughes, L. M., Bao, J., Hu, Z. L., Honavar, V. and Reecy, J. M. 2008. Animal trait ontology: The importance and usefulness of a unified trait vocabulary for animal species. *J. Anim. Sci.* **86**, 1485-1491.
- Jiang, Z., Michal J. J., Chen, J., Daniels, T. F., Kunej, T., Garcia, M. D., Gaskins, C. T., Busboom, J. R., Alexander, L. J. and Wright, R. W. J. 2009. Discovery of novel genetic networks associated with 19 economically important traits in beef cattle. *Int. J. Biol. Sci.* **5**, 528-542
- Kang, B. and Kim, H. 2011. Web-based ontology edition operation system. PCT/KR2011/000595
- Kim, H., Shim, J. E., Shin, J. and Lee, I. 2015. EcoliNet: a database of cofunctional gene network for Escherichia coli. *Database*, bav001.
- Kim, N. K., Lim, D., Lee, S. H., Cho, Y. M., Park, E. W., Lee, C. S. and Yoon, D. 2011. Heat shock protein B1 and its regulator genes are negatively correlated with intramuscular fat content in the Longissimus Thoracis muscle of Hanwoo (Korean cattle) steers. *J. Agric. Food Chem.* **59**, 5657-5664.
- Lim, D., Lee, S. H., Kim, N. K., Cho, Y. M., Chai, H. H., Seong, H. H. and Kim, H. 2013. Gene Co-expression Analysis to Characterize Genes Related to Marbling Trait in Hanwoo (Korean) Cattle. *Asian Australas. J. Anim. Sci.* **26**, 19.
- Lim, D., Kim, N. K., Park, H. S., Lee, S. H., Cho, Y. M., Oh, S. J. and Kim, H. 2011. Identification of candidate genes related to bovine marbling using protein-protein interaction networks. *Int. J. Biol. Sci.* **7**, 992-1002.
- Obayashi, T., Hayashi, S., Shibaoka, M., Saeki, M., Ohta, H. and Kinoshita, K. 2008. COXPRESdb: a database of co-expressed gene networks in mammals. *Nucleic Acids Res.* **36**, 77-82.
- Peri, S., Navarro, J. D., Kristiansen, T. Z., Amanchy, R., Surendranath, V., Muthusamy, B. and Rashmi, B. P. 2004. Human protein reference database as a discovery resource for proteomics. *Nucleic Acids Res.* **32**, 497-501.
- Peter, L. and Steve, H. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559.
- Reverter, A., Hudson, N., Wang, Y., Tan, S., Barris, W., Byrne, K., McWilliam, S., Bottema, C., Kister, A. and Greenwood, P. 2006. A gene coexpression network for bovine skeletal muscle inferred from microarray data. *Physiol. Genomics* **28**, 76.
- Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W. and Leontis, N. 2007. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* **25**, 1251-1255.
- Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F. H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A. and Koeppen, S. 2005. A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome. *Cell* **122**, 957-968.
- Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P. and Jensen, L. J. 2010. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **39**, 561-568.

초록 : 소의 경제형질 관련 유전자 네트워크 분석 시스템 구축

임다정¹ · 김형용² · 조용민¹ · 채한화¹ · 박종은¹ · 임규상¹ · 이승수^{1*}

(¹농촌진흥청 국립축산과학원, ²인실리코젠)

가축의 경제형질은 대부분 복합형질 상태이며, 많은 유전자와 생물대사회로에 의해 조절된다. 시스템 생물학은 생명현상을 하나의 복합체로 가정하고, 형질에 관여하는 유전자들에 대한 기능적 관계를 분석하는 학문이다. 유전자 네트워크는 시스템 생물학의 하나의 연구분야로써, 유전자 기능의 상관관계를 지도화하여 오믹스 데이터를 통합 분석하여 해석한다. 유전자 네트워크는 단백질-단백질 상호작용, 공발현, 조절인자, 유전자형 기반으로 다양한 유전자의 기능적 상호작용을 표현할 수 있다. 또한, 네트워크를 구성하기 위해서는 유전자 간 연결 정도에 가중치를 두거나, 인접한 유전자 수 계산 등의 네트워크 토폴로지 알고리즘이 적용된다. 가축에서는 이러한 연구가 단형질에 대한 유전자 발현, 단백질 상호작용 등에 국한되어 있는 실정이다. 본 논문에서는 유전자 공발현 네트워크와 단백질-단백질 상호작용 네트워크 분석법을 확립하고 소의 102개 경제형질에 대하여 유전자 네트워크 분석 결과에 대한 데이터베이스를 구축하였다. 102개의 경제형질은 Animal Trait Ontology (ATO) 명명법에 의하여 분류하여 제공하였다. 각 형질에 포함된 유전자 리스트는 Animal QTL database에서 제공하는 양적유전형질좌위의 물리적 위치에 존재하는 유전자군을 추출하였다. 유전자 공발현 네트워크는 R의 WGCNA 패키지를 활용하였으며, 단백질-단백질 상호작용 네트워크는 Human Protein Reference Database에서 사람과 소의 orthologous group에 포함된 유전자를 대상으로 단백질 상호작용 관계를 규명하였다. 네트워크 분석 결과는 관계형 테이블로 구축하였으며, 구축한 데이터베이스를 관련 연구진에게 공유하기 위하여 웹 기반의 유전자 네트워크 가시화 시스템을 구현하였다(<http://www.nabc.go.kr/cg>). 웹 데이터베이스 구현을 위하여 Ontle 프로그램을 활용하여 다양한 방식으로 유전자 네트워크 가시화 작업을 수행하였다. 이 시스템을 통하여 사용자는 관련 형질의 후보 유전자군 탐색, 유전자 네트워크 분석 결과, 유전자 사이의 기능적 연결관계를 손쉽게 살펴볼 수 있게 될 것이다.