

닭 유전체 연구의 최근 동향

서성원¹ · 백운기² · 이준현^{1,†}

¹충남대학교 동물자원과학부, ²국립중앙과학관

Recent Status of Chicken Genome Researches

S. W. Seo¹, W. K. Paek² and J. H. Lee^{1,†}

¹Division of Animal Science and Resources, Chungnam National University

²National Science Museum

ABSTRACT Chicken has been being used as the protein sources for humans for long times. Since the release of chicken genome sequencing, large efforts have been carried out for identifying valuable genomic information in chicken including the genes affecting quantitative traits, development and immune responses. In this review, the recent progress of genomic researches has been discussed including the available world wide genetic materials in chicken.

(Key words : chicken, genome sequencing, genetic materials)

서 론

닭(*Gallus gallus domesticus*)은 인류에게 필요한 단백질을 손쉽게 공급해 주는 단백질 공급원으로 중요한 역할을 담당하여 오고 있다. 2004년 닭의 genome sequencing이 발표된 이후 닭은 유전체 연구의 중심에 서게 되었으며, 특히 발생학과 척추동물의 모델 동물로서 그 중요성이 더욱 증대되고 있다. 본 논문은 최근 닭의 유전체 연구를 조명하여 봄으로써 닭을 연구하는 연구자들에게 보다 많은 정보를 제공하기 위하여 쓰여졌다.

본 론

1. 닭의 Genome Sequencing 현황 및 응용

닭의 genome sequencing은 인간 genome sequencing을 통해 축적된 경험과 BAC(Bacterial Artificial Chromosome) library와 같이 잘 만들어진 많은 유전 자원들 덕분에 인간 genome sequencing에 비하여 상대적으로 적은 노력과 비용이 소요되었다(Hillier 등, 2004). 닭은 포유동물과 비교하여 30% 정도 밖에 되지 않는 적은 유전체의 크기를 가지고 있어 노력에

비하여 양질의 sequencing 결과를 얻을 수 있었다. 또한, sequence 분석에 있어서 포유동물의 40~50%를 차지하는 반복 염기서열이 닭에서는 11%밖에 존재하지 않기 때문에 이 역시 양질의 sequence를 얻을 수 있는 바탕이 마련되었다. 이런 반복 염기 서열은 주로 centromere 부분에 존재하며 인간의 genome sequence의 초안이 공개된 후 7년이 지난 지금까지도 완벽하게 인간의 sequence를 분석하지 못한 이유가 이 반복 염기 서열에 있다.

닭의 genome sequencing에 사용된 개체는 야계(jungle fowl) 암컷(ZW)으로(Fumihito 등, 1994) 전체 genome의 약 6.6X를 포함하는 첫 번째 초안이 2004년에 발표되었다(Hillier 등, 2004). 첫 번째 초안의 경우, 최종 sequence 분석 결과 약 98%의 sequence가 분석되었고, 5~10%의 유전자가 아직도 정확히 분석되지 않은 것으로 추정되었다. 이런 분석을 통하여 닭이 가지고 있는 유전자 수는 약 2만에서 2만 3천개가 있다고 추정되었으며, 아직도 3,000여개의 부정확한 유전자의 수는 유전자의 추정 및 보완 실험에 대하여 많은 연구와 노력이 필요하다라는 것을 의미하기도 한다(Eyras 등, 2005). 또한, 닭과 인간 유전자 비교를 통하여 단지 37개의 유전자만이 닭 유전체에만 특이적으로 존재하는 것으로 추정되어(Castelo 등, 2005), 분류학적으로 상당히 먼 포유동물(인간)과 조류(닭)에

[†] To whom correspondence should be addressed : junheon@cnu.ac.kr

서 기능을 가진 유전자는 보존이 매우 잘 되어 있다는 것을 알려주고 있다(Hillier 등, 2004).

2006년 5월에 미국 세인트루이스에 있는 Washington University School of Medicine의 genome sequencing center를 주축으로 새로운 초안(assembly build 2.1)이 발표되었다. 닭 유전체의 상염색체 중 1~28번과 32번, 성염색체(W, Z), 원래의 연관그룹(LG: Linkage Group)에 두 개의 연관그룹(LGE64와 LGE22C19W28_E50C23)이 추가로 assemble 되었다. 19개 Ultracontig의 N50은 15.5 Mb이며, 가장 큰 ultracontig는 3번 염색체에 위치하며, 80.3 Mb의 크기를 가지고 있다. 26개 supercontig의 N50은 11.1 Mb이고, 가장 큰 supercontig는 52 Mb로 2번 염색체에 위치해 있다. 그 밖의 contig는 N50이 45 kb인 5863개의 작은 contig들로 그 중 가장 긴 contig는 625 kb이다. 여기서 N50이라는 것은 생물정보학에서 contig의 크기를 나타내는 기준으로 50%의 genome sequence가 N50 contig 혹은 그보다 더 큰 contig에 존재함을 말한다. 다시 말하면, 19개의 ultracontig를 크기의 순서대로 나열한 후에 큰 순서대로 차례로 sequence의 수를 더하는 과정 중 어떤 contig의 크기를 더했을 때 그 때까지의 합이 ultracontig 19개의 크기를 모두 더한 수의 50% 이상이 된다면, 그 때의 contig 크기를 N50라 한다. Assembly란 contig를 순서에 따라 계속 연결하는 과정을 말하며, contig 간에는 sequence gap이 있기 때문에 assemble된 contig의 N50 수치가 클수록 assembly의 완성도가 높다고 할 수 있다.

닭은 인류에게 단백질의 공급원으로서 매우 중요한 위치를 차지하고 있기 때문에 닭의 유전체의 연구는 농업에 있어 그 가치가 매우 높으며 최근 비교유전체학, 진화생물학, 발생학적인 측면에서도 그 가치가 높아지고 있다. 닭 유전체를 sequencing한 이유 중 하나는 척추동물 중 조류를 sequencing함으로써 비교유전체학의 이해를 높이며 궁극적으로 인간 유전체 해석을 정확히 하는데 그 의미가 있다(Hillier 등, 2004). 최근 연구에 의하며 닭과 인간 간에 ultra-conserved regions(UCR)이 400개가 넘게 밝혀졌다(Sandelin 등, 2004). 이들 UCR중 기능을 가진 유전자(gene)와 상관없는 sequence들이 많아 종의 진화를 해석하는데 좋은 자료가 되고 있다(Ovcharenko 등, 2005). 또한, genomic imprinting과 같은 부모의 특이한 유전자 발현은 조류에서는 밝혀져 있지 않으며, 포유동물에서만 밝혀져 있다. 따라서 닭이 imprinting의 기원에 대하여 단서를 제공할 뿐만 아니라 조류에서 포유동물로 진화될 시 유전자의 선발과 관련된 중요한 정보를 제공해 줄 수 있다는 가능성을 제시하고 있다.

닭은 포유류와는 달리 발생이 알에서 이루어지기 때문에

배이를 쉽게 얻을 수 있으며 조작이 쉽기 때문에 발생학을 연구하는 학자들에게 매우 중요한 연구 재료가 되어 왔다(Burt, 2004; Stern, 2005). 최근 유전자 조작 기술이 닭에게 적용됨으로써 닭이 유전자의 기능을 연구하기 위한 적합한 모델 동물로 생각되고 있다.

닭의 genome sequencing이 끝남과 동시에 닭에 존재하는 SNP(single nucleotide polymorphism)를 분석하기 위한 협의체가 구성되었으며(Wong 등, 2004; Wang 등, 2005; ChickVD), 이 협의체는 육계, 산란계, 적색야계, Silkie 종간의 비교 분석을 통하여 2백 8십만 개의 SNP를 찾아냈으며, 닭은 nucleotide 변이가 인간에 비하여 약 6배가 빠른 것으로 나타났다(Ellegren, 2005). 밝혀진 SNP의 정확도를 높이기 위하여 실시한 두 번의 sequencing 결과 전체 SNP의 94%가 재확인되었다. 이런 SNP의 연구는 경제적으로 중요한 의미를 가지는 양적 형질 유전자 좌위(QTL: quantitative trait loci)의 원인 SNP를 찾는 데 이용이 가능하며, 닭은 다른 동물과 달리 기준 집단(resource population)을 쉽게 만들 수 있다는 장점을 가지고 있어 앞으로 닭을 이용한 연구의 속도는 매우 빨리 진행될 것으로 예측된다.

2. 닭유전체 연구 자원 현황

닭은 산란을 위하여 개량된 산란계(layer)와 고기를 목적으로 개량된 육용계(broiler)로 나뉘어 선발되어 왔다. 닭은 앞서도 언급하였듯이 포유동물과 조류의 진화를 풀어나갈 수 있는 모델 동물임과 동시에 발생학적으로 매우 중요한 역할을 하고 있다. 또한, 단백질의 공급원으로 그 역할을 충분히 수행하고 있다. 1990년대를 기점으로 발달된 구조유전체학의 도움으로 경제 형질 관련 유용 유전자를 탐색하여 선발에 이용하려는 노력이 현재까지 계속되어오고 있으며, 이를 위한 연구 자원(genetic resources)이 많이 만들어졌다.

유전체 연구를 위한 가장 중요한 재료중 하나가 기준 집단(reference population)인데, 닭의 기준 집단은 연관 유전자 지도 작성을 목적으로 1990년대 초반에 Compton(CT; Bumstrad and Palyga, 1992)과 East Lansing(EL; Crittenden 등, 1993) 집단이 만들어졌다. 이 두 집단은 international reference population이라고 불렸으며, 세계의 어떤 연구자에게도 재료의 제공되어 국제적인 협력에 중요한 역할을 담당하였으며, 많은 결과를 빠른 시간에 얻을 수 있는 기회를 제공하였다. 그 후 양적 형질 유전자 좌위(QTL)를 찾기 위한 많은 기준 집단이 제작되었는데(Abasht 등, 2006), 주로 이 기준 집단들은 F₂, 퇴교배(backcross, BC)와 F₁으로 구성되어 있다. 그 다음 세대인 두 번째 세대는(G₂) 부모중 한 집단과의 퇴교배로 만들

어지거나(BC design), 제1세대의 개체간 교배로 이루어진다(F₂ design).

닭의 genome sequencing의 결과로 얻어진 자료중 하나가 SNP의 확보이며, 그 후 지금까지 SNP discovery에 대한 광범위한 연구가 시도되어 오고 있다. 현재 dbSNP에 저장된 닭의 SNP수는 3,335,290개로 이는 검증되지 않은 SNP로 앞으로 무수히 많은 재확인 작업을 거쳐 유용하게 이용될 수 있을 것으로 사료된다(<http://www.ncbi.nlm.nih.gov>).

2000년대 들어오면서 많은 유전자의 발현을 한 번에 관찰하려는 microarray 관련 연구가 시작이 되었다. 초기의 microarray는 cDNA microarray로서 유럽을 중심으로 구성된 모임인 ARK-genomics를 통하여 만들어졌는데, 면역기관에서 추출한 mRNA를 이용하여 cDNA clone을 이용하여 만든 5K Immune cDNA microarray가 있으며, 크기는 13,000개의 cDNA를 함유한 microarray가 제작되었다. 최근 cDNA microarray보다 연구의 정확도와 재연성이 높은 oligomicroarray를 이용하려는 시도가 계속되고 있으며, microarray를 전문적으로 만들어내는 회사가 array 제작을 주도하고 있다. 그 예로 Affymetrix 사는 2만 8천개 이상 유전자를 포함한 chicken genome array를 제작하였으며, gilent 사는 chicken whole genome 44K array를 제작하여 연구에 이용하고 있다(Masker 등, 2007; <http://tard.state.tx.us/>).

3. 닭유전체의 유용 유전자 발굴 현황

타 가축과 마찬가지로 현재 인간과 생쥐를 제외한 동물의 유전양상은 호주의 Frank Nicholas 박사가 운영하고 있는 Online Mendelian Inheritance in Animals(OMIA: <http://omia.angis.org.au/>)에 잘 요약되어 있다. 닭에서 현재까지 밝혀진 유전 양상의 요약은 Table 1에 나타나 있으며, 이중 22개의 형질만이 분자 수준에서 원인 유전자가 밝혀져, 앞으로 연구해야 할 내용이 많음을 의미한다.

닭은 생산성을 높이기 위하여 많은 개량이 되어 왔는데 이 과정에서 유전자의 다형성이 감소되었으며, 이 유전자의 다형성 감소가 질병에 관한 저항성을 낮추었다고 생각되고 있다. 따라서 최근의 연구는 질병의 저항성을 높이는 방안을

제시하기 위하여 많은 노력을 하고 있다. 현재까지 닭의 질병 중 가장 연구가 많이 된 질병은 Marek disease로 지난 30년간 꾸준히 연구가 되어 왔다. 미국 농무성(USDA)이 Marek's disease(MD)에 감수성과 저항성 계통을 육성하여 원인이 되는 유전자를 검색하고 있으며, 최근 Growth hormone이 MD virus와 상호작용하여 질병 저항성을 나타낸다는 보고가 있다(Lui 등, 2001). 또한, 최근 연구 결과에 의하면 MD virus가 주조직 적합성 복합체(MHC: Major Histocompatibility Complex)의 class II 유전자들의 발현을 증가시킨다는 보고가 나와 질병의 저항성 및 감수성이 MHC 유전자와 밀접한 관계가 있음을 나타내고 있다(Niikura 등, 2007). 그러나 아직까지 질병 저항성 및 감수성과 관련된 유전자에 대한 연구 결과는 미흡한 실정이며, 한국에서도 최근 매년 발생하는 조류독감의 피해가 증가함에 따라 질병 저항성에 관한 지속적인 관심과 연구를 집중하고 있다.

2000년대에 들어서면서부터 육계의 성장이나 산란계의 산란율과 같이 경제 형질이라고 알려진 양적 형질을 개량하는데 분자유전학적 접근 방법이 이용되고 있다. 닭의 경우 Chicken Quantitative Trait Loci database(ChickenQTLdb; <http://www.animalgenome.org/QTLdb/chicken.html>)에 그 최근까지의 연구 결과가 요약되어 나타나 있는데, 지난 10년 동안 닭에서 112개의 형질에 대하여 657개의 양적 형질 유전자 좌위가 밝혀졌다. 이와 같이 닭에 있어서 양적 형질과 관련이 된 염색체상의 부분이 밝혀졌으나 아직까지 원인 유전자와 원인 변이를 찾아 상업적으로 이용이 되는 마커는 전무한 실정이다. 따라서 앞으로의 연구는 이런 변이를 찾는 데 주력할 것으로 사료된다.

최근에 international consortium을 통해 닭 유전체와 기타 몇몇 조류(칠면조, 콘돌 등)의 유전 정보를 통합적으로 보유하는 데이터베이스가 구축 중에 있다(birdbase; <http://birdbase.net>). Birdbase는 Gallus GBrowse를 통해 닭의 유전자, EST, SNP, Marker 등의 정보 및 microarray 연구 결과를 통합적으로 보유하고 있으며, Illumina G2 차세대 sequencer를 이용한 sequence 분석 결과도 제공할 예정이다. 델라웨어 주립대학(University of Delaware)의 Schmidt 교수팀을 주축으로 구성된 Reactome project(Matthews 등, 2009)를 통해 현재까지 사람의 총 2,734개의 생화학 반응 중 1,838개가 닭의 유전체에 존재함이 밝혀졌다. 이후 계속된 노력을 통해 더 많은 대사 경로가 닭의 유전체에서 밝혀질 것으로 예상된다. 닭은 다른 모델 생물에 비하여 포유류와 비교적 최근에 분화된 종으로 가장 양질의 유전체 assembly를 가지고 있는 동물이기 때문에, 닭 유전체의 대사관련 유전자에 관한 연구는 포유동물의

Table 1. OMIA summary of inherited disorders and traits in chicken (March, 2009)

Total phenos	Single-locus phenos	Phenos characterized at the molecular level	Potential model for human disease
179	72	22	38

대사 조절 기작을 밝히는 데에 매우 중요한 자료로 쓰일 수 있다. 특히, 포유동물 유전자의 생성, 소멸 시기를 밝히기 위한 phylogenetic foot-printing에 닭의 유전자가 많이 사용되어지고 있다. 따라서 닭 유전체에 관한 연구는 진화의 원리를 이해하는데 매우 중요한 정보를 제공할 수 있을 것으로 기대된다.

4. 국내 가금 유전체 연구와 앞으로의 과제

가금 분야에서 유전체를 이용한 연구는 2002년 농촌진흥청 주관 바이오그린 21사업의 시작으로 다양한 연구가 시작되었다. 이 연구의 결과 형질 전환 닭 생산, 한국재래닭의 유전체 분석, 한국재래닭 경제 형질 관련 QTL 분석, 프로테오믹스를 이용한 한국재래닭의 단백질체 분석, 한국재래닭의 대량 유전자 기능 분석 연구 등이 이루어졌으나, 소와 돼지의 연구에 비교하여 보면 투자된 연구비나 연구원의 수가 매우 적었음을 알 수 있다. 특히 이런 연구를 통하여 한국재래닭에 대한 의미를 재조명해 봄과 동시에 유전자원의 중요성을 국내외에 알린 계기가 되었다. 그러나 국내 가금 유전체의 연구가 외국의 선진 유전체 연구와 비교하여 손색이 없을 정도로 우수한 연구가 되기 위해서는 체계적인 연구 협조와 연구의 바탕이 될 수 있는 가금 유전 재료 확보가 매우 시급하다.

결 론

전통적인 육종 방법을 통하여온 현재까지의 개량은 분자 유전학적 marker의 추가로 인하여 개량의 속도가 가속화되고 있다. 본 논문의 chicken genome sequencing을 통하여 나온 결과가 현재까지 어떤 연구 및 응용 분야가 가능한지 살펴보았다. 특히 가금은 앞에서 설명했듯이 척추동물의 기본 모델이 될 뿐 아니라 발생학적으로 매우 중요한 모델 동물로 자리잡아가고 있다. BAC library를 포함한 전 세계에 가용한 유전재료들의 발달이 닭을 모델 동물로 이용한 연구를 가속화시킬 것으로 생각된다.

적 요

닭은 인간에게 계란과 닭고기를 공급하는 중요한 동물임과 동시에 발생학적, 비교 유전학적으로 매우 중요한 위치를 차지하고 있다. 2004년 발표된 닭의 genome sequencing이후

닭의 유전체 연구에 많은 변화가 일어났다. 본 논문은 그동안 닭의 유전체 연구에 사용된 기초 집단 등 닭 유전체 연구의 재료 현황 및 연구 현황을 재조명하여 봄으로서 앞으로 닭 유전체 연구의 방향을 설정하는데 도움을 주고자 작성하였다.

사 사

본 연구는 농촌진흥청의 바이오그린 21사업 연구동향분석과제(동물의 유용유전자 탐색, 개발 및 이용기술 확립 연구)의 지원에 의해 수행되었습니다.

인용문헌

- Abasht B, Dekkers JC, Lamont SJ 2006 Review of quantitative trait loci identified in the chicken. *Poult Sci* 85(12):2079-2096.
- Bumstead N, Palyga J 1992 A preliminary linkage map of the chicken genome. *Genomics* 13:690-697.
- Burt DW 2004 The chicken genome and the developmental biologist. *Mech Dev* 121:1129-1135.
- Castelo R, Reymond A, Wyss C, Camara F, Parra G, Antonarakis SE, Guigó R, Eyra E 2005 Comparative gene finding in chicken indicates that we are closing in on the set of multi-exonic widely expressed human genes. *Nucleic Acids Res* 33:1935-1939.
- Crittenden LB, Provencher L, Santangelo L, Levin I, Abplanalp H, Briles WE, Dodgson JB 1993 Characterization of a red jungle fowl by White Leghorn backcross reference population for molecular mapping of the chicken genome. *Poult Sci* 72:334-348.
- Ellegren H 2005 The avian genome uncovered. *Trends Ecol Evol* 20:180-186.
- Eyra E, Reymond A, Castelo R, Bye JM, Camara F, Flicek P, Huckle EJ, Parra G, Shteynberg DD, Wyss C, Rogers J, Antonarakis SE, Birney E, Guigo R, Brent MR 2005 Gene finding in the chicken genome. *BMC Bioinformatics* 6:131.
- Fumihito A, Miyake T, Sumi S, Takada M, Ohno S, Kondo N 1994 One subspecies of the red junglefowl (*Gallus gallus gallus*) suffices as the matriarchic ancestor of all domestic

- breeds. *Proc Natl Acad Sci USA* 91:12505-12509.
- International Chicken Genome Sequencing Consortium 2004 Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695-716.
- Liu H, Kung H, Fulton JE, Morgan RW, Cheng HH 2001 Growth hormone interacts with the Marek's disease virus SORF2 protein and is associated with disease resistance in chicken. *Proc Natl Acad Sci USA* 98:9203-9208.
- Masker K, Golden A, Gaffney CJ, Mazack V, Schwindinger WF, Zhang W, Wang LH, Carey DJ, Sudon M 2007 Transcriptional profile of rous sarcoma virus transformed chicken embryo fibroblasts reveals new signaling targets of viral-src. *Virology* 364(1):10-20.
- Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, Garapati P, Hemish J, Hermjakob H, Jassal B, Kanapin A, Lewis S, Mahajan S, May B, Schmidt E, Vastrik I, Wu G, Birney E, Stein L, D'Eustachio P 2009 Reactome knowledgebase of biological pathways and processes. *Nucleic Acids Res* 37:D619-622.
- Niikura M, Kim T, Hunt HD, Burnside J, Morgan RW, Dodgson JB, Cheng HH 2007 Marek's disease virus up-regulates major histocompatibility complex class II cell surface expression in infected cells. *Virology* 359(1):212-219.
- Ovcharenko I, Loots GG, Nobrega MA, Hardison RC, Miller W, Stubbs L 2005 Evolution and functional classification of vertebrate gene deserts. *Genome Res* 15:137-145.
- Sandelin A, Bailey P, Bruce S, Engström PG, Klos JM, Wasserman WW, Ericson J, Lenhard B 2004 Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. *BMC Genomics* 5:99.
- Stern CD 2004 The chick embryo-Past, present and future as a model system in developmental biology. *Mech Dev* 121: 1011-1013.
- Wang J, He X, Ruan J, Dai M, Chen J, Zhang Y, Hu Y, Ye C, Li S, Cong L, Fang L, Liu B, Li S, Wang J, Burt DW, Wong GK, Yu J, Yang H, Wang J 2005 Chick VD: A sequence variation database in the chicken genome. *Nucleic Acids Res* 33:D438-D441.
- Wong GK, Liu B, Wang J, Zhang Y, Yang X, Zhang Z, Meng Q, Zhou J, Li D, Zhang J, Ni P, Li S, Ran L, Li H, Zhang J, Li R, Li S, Zheng H, Lin W, Li G, Wang X, Zhao W, Li J, Ye C, Dai M, Ruan J, Zhou Y, Li Y, He X, Zhang Y, Wang J, Huang X, Tong W, Chen J, Ye J, Chen C, Wei N, Li G, Dong L, Lan F, Sun Y, Zhang Z, Yang Z, Yu Y, Huang Y, He D, Xi Y, Wei D, Qi Q, Li W, Shi J, Wang M, Xie F, Wang J, Zhang X, Wang P, Zhao Y, Li N, Yang N, Dong W, Hu S, Zeng C, Zheng W, Hao B, Hillier LW, Yang SP, Warren WC, Wilson RK, Brandström M, Ellegren H, Crooijmans RP, van der Poel JJ, Bovenhuis H, Groenen MA, Ovcharenko I, Gordon L, Stubbs L, Lucas S, Glavina T, Aerts A, Kaiser P, Rothwell L, Young JR, Rogers S, Walker BA, van Hateren A, Kaufman J, Bumstead N, Lamont SJ, Zhou H, Hocking PM, Morrice D, de Koning DJ, Law A, Bartley N, Burt DW, Hunt H, Cheng HH, Gunnarsson U, Wahlberg P, Andersson L, Kindlund E, Tammi MT, Andersson B, Webber C, Ponting CP, Overton IM, Boardman PE, Tang H, Hubbard SJ, Wilson SA, Yu J, Wang J, Yang H, International Chicken Polymorphism Map Consortium 2004 A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature* 432:717-722.

(접수: 2009. 3. 17, 수정: 2009. 5. 7, 채택: 2009. 6. 6)