The investigation of novel SNPs in Taiwan Country Chicken using bioinformatics tools

E.-C.Lin (林恩仲) and Y. H. Wang (王怡惠)
Department of Animal Science and Technology
National Taiwan University (臺灣大學動物科學技術學系)

Single nucleotide polymorphism (SNP) is the basic and common variation in genomic sequence. The size of chicken genome is around 1.2×10^9 bps with 3.28×10^6 SNPs discovered. Majority of those SNPs were detected by the international cooperative chicken genome project using genomic DNA of broiler, layer, silky and red jungle fowl. Another source for detecting SNPs (esp., in exons) is from Expressed sequence tag (EST) libraries. Some of SNPs happened in exons change the corresponding amino acids in the related protein sequence, which are so called non-synonymous SNPs.

參考文獻

- **Amaya, E., Offield, M.F., and Grainger, R.M. (1999). Frog genetics: Xenopus tropicalis jumps into the future. Trends Genet *14(7)*:253-5.**
- **Balch, G.C., Velez-Espino, L.A., Sweet, C., Alaee, M., and Metcalfe, C.D. (2006). Inhibition of metamorphosis in tadpoles of Xenopus laevis exposed to polybrominated diphenyl ethers (PBDEs). Chemosphere *64*, 328-338.**
- **Belmonte, J.C., Ellis, J., Hochedlinger, K., and Yamanaka, S. (2009). Induced pluripotent stem cells and reprogramming: seeing the science through the hype. Nat Rev Genet *10(12)*:878-83.**
- **Dale, L., and Slack, J. M. (1987). Fate map for the 32-cell stage of Xenopus laevis, Development *99*, 527-51.**
- **Gilbert, S.F. (2003). Developmental Biology 7th edition. Sinauer Associates Inc. pg. 81-87; 305-337.**
- **Gurdon, J.B. (1962). The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. J Embryol Exp Morphol *10*:622-40.**
- **Harland, R., and Gerhart, J. (1997). Formation and function of Spemann's organizer. Annu Rev Cell Dev Biol *13*, 611-67.**
- **Heasman, J. (2002). Morpholino oligos: making sense of antisense? Dev Biol *243*, 209-14.**
- **Khokha, M.K., Chung, C., Bustamante E.L., Gaw, L.W., Trott, K.A., Yeh, J., Lim, N., Lin, J.C., Taverner, N., Amaya, E., Papalopulu, N., Smith, J.C., Zorn, A.M., Harland, R.M., and Grammer, T.C. (2002). Techniques and probes for the study of Xenopus tropicalis development. Dev Dyn *225(4)*:499-510.**
- **Kloas, W. (2002). Amphibians as a model for the study of endocrine disruptors. Int Rev Cytol *216*, 1-57.**
- **Kroll, K. L., and Amaya, E. (1996). Transgenic Xenopus embryos from sperm nuclear transplantations reveal FGF signaling requirements during gastrulation, Development *122*, 3173-83.**

- **Lutz, I., and Kloas, W. (1999). Amphibians as a model to study endocrine disruptors: I. Environmental pollution and estrogen receptor binding. Sci Total Environ *225*, 49-57.**
- **Nieuwkoop, P.D., and Faber, J. (1994). Normal table of Xenopus laevis (daudin). Garland Publishing. Inc., New York and London.**
- **O'Rourke, D.P. (2007). Amphibians used in research and teaching. ILAR J *48*, 183-187.**
- **Sive, H. L. (2000). Early Development of *Xenopus laevis*: A Laboratory Manual. Cold Spring Harbor Laboratory Press.**
- **Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126(4)*:663-76.**
- **Tseng, H.T., Shah, R., and Jamrich, M. (2004). Function and regulation of *FoxF1* during *Xenopus* gut development. Development *131(15)*:3637-47.**
- **Wallingford, J. B. (1999). Tumors in tadpoles: the Xenopus embryo as a model system for the study of tumorigenesis, Trends Genet *15*, 385-8.**
- **Zhang, J., Rosenthal, A., de Sauvage, F. J., and Shivdasani, R. A. (2001). Downregulation of Hedgehog signaling is required for organogenesis of the small intestine in Xenopus, Dev Biol *229*, 188-202.**
- **Zhang, L., El-Hodiri, H. M., Ma, H. F., Zhang, X., Servetnick, M., Wensel, T. G., and Jamrich, M. (2003). Targeted expression of the dominant-negative FGFR4a in the eye using Xrx1A regulatory sequences interferes with normal retinal development, Development *130*, 4177-86.**

The Taiwan Country chickens are the native breeds in Taiwan. There are 10 lines (5 male and 5 female) long-term selected with different criteria from the same ancestral population in National Chung-Hsin University.The Taiwan Country chickens are the most important meat-type breed with more than 60% market share in the local poultry market. It is important to maintain those purebred lines for obtaining higherheterosis and production efficiency. The information of genetic polymorphism is needed to achieve such goals in the local poultry industry as well as the scientific research. This study was designed to identify SNPs from several EST libraries of two major selection lines in National ChungHsing University.

There were 4~6 hens randomly selected from the Taiwan Country chicken lines for the EST clone library construction: the male line B selected on body weight and comb areaat10~12 weeks of age; the femalelineL2 selected on egg production at 40 weeks of age (Figure 1). Six egg-laying related tissues (pituitary, liver, adipose, ovary, oviduct, and shell gland) and muscle tissue from those hens during highly egg- laying period were used. Totally, 42,404 EST clones from the 7 libraries were randomly selected for DNA sequencing using Applied Biosystems 3130 DNA Analyzer.

Figure 1. The Taiwan Country chicken selection lines of B (high meat production, left) and L2 (high egg production, right). (Picture obtained from: http://www.angrin.tlri.gov.tw/taiwan/chicken/chicken.html)

Those EST sequences were analyzed by using Phred programfor base-calling of all the bases shown in the sequencing chromatograph and clean-up of those low quality (QV≦ 30) bases and Phrap program to assemble those high quality EST sequences. Those assembledcontigs obtained from Phrap were annotated and localized by Blastn against NCBInucleotide and genome databases. Those contigscontaining at least 5 ESTs were further predicted the possibility of any particular SNP happened≧ 15% ofthose overlapped ESTs by PolyPhred program. Those contig sequences with SNP predicted were searched against NCBI dbSNPdatabase using Blastn to make sure whether those SNPs found by PolyPhred are novel (not found in dbSNP) or known (matchedto at least one SNP in dbSNP).

There were 42,404 EST clones randomly selected for sequencing and 36,463 high quality sequences left after trimming those low quality bases by Phred. The total numbers of assembled contigs, sequences not qualifiedly to other ESTs (Singlets), and sequences with problem during alignment and assembling were 3,455, 11,649, and 1,840, respectively, from those high quality sequences in the seven Taiwan Country chicken EST libraries using Phrap.The total number of SNPs predicted by PolyPhred was 1,107 with only 297 (26.8%) known and 810 (73.2%) still not found in NCBI dbSNP. With restriction of a contig with at least 5 ESTs and a SNP happened ≧ 15% of those overlapped ESTs in that contig, those SNPs predicted of very high Polyphred score (98~99) and moderately high (95~97) were 951 and 56 (91% of total SNPs predicted). The numbers of SNPs predicted detected on different location of a particular gene were60, 704, 10, 121, 5, and 207accordingly on 5'UTR, exon, intron, 3'UTR, pseudogene, and unknown position. Among the 704 SNPs found in exons, there were 546 and 158 detected as synonymous and non-synonymous types, respectively.

The blood samples of 94 hens were collected equally from lines B and L2. The genomic DNA was extracted from those blood samples in our lab. Among those predicted SNPs with PolyPhred score ≧ 98, there were 84 predictedSNPs and 1 negative control(not detected with any SNP) sequences selected for further validation using GenomeLab™SNPstream® SNP (Backman Coulter, CA) genotyping in the Microarray Core of NTU 基因體醫學研究中心.The 85 predicted SNP sequences of high PolyPhred score were validated from 94 hens of lines B and L2. We found that 60 of true positive (70.6%), 24 of possible false positive (28.2%), and 1 of true negative (the negative control sequence, 1.2%).After statistical comparison between the genotypic frequencies in the lines B and L2 by chi-square test, there were 34 out of the 60 true positivensSNPs located on 29 different genes found with significant difference (p-value < 0.05). Those genes have certain biological functions, such as growth of egg yolk, immune protection, formation of egg shell, metabolism, regulatory, development, etc. (Table 1)

SNP prediction from EST sequences (or so called transcriptome sequences) has been considered not as good as from genomic sequences in terms of quality. But the cost of obtaining abundant EST sequences is much cheaper than genome mapping project. In addition, EST library might be applied to those particular breeds and selection lines different from commercial and standard experimental ones. These results clearly showed that the quality procedure of base-calling with higher QV, assembling contig and predicted SNPs with certain limitations might be greatly helpful for the total quality of prediction. Those high quality SNPs predicted from different breeds and lines will lead to a lot of important applicationssuch as phylogenetic analysis, domestication, functional studies, marker-assisted selection programs, etc.

This study was cooperated by Dept. of Animal Science and Technology (E.-C. Lin, Y. H. Wang,C. X. Wang, and S. T. Ding),Institute Biochemistry and Molecular Biology(J. J. Chen and S. C. Lu), National Taiwan Univ.; Dept. of Animal Science (C. F. Chen, M. C. Huangand Y. P. Lee), Institute of Microbiology and Public Health (S.-H. Chiou), National Chung-Hsing University;National Center for High-Performance Computing (C. W. Yeh andC. H. Hsieh); Institute of Molecular Medicine, National ChengKung University (H. S.Sun);Dept. of Animal Science and Biotechnology, Tunghai University (B. R. Ou and W. T. K. Cheng).

2011.10.12 林恩仲老師醫學院演講

Table 1. The genes and related biological functions found with nsSNP significantly distributed between hens from the selection lines of B and L2.

Gene found with nsSNP significantly distributed between lines B and L2	Biological Function
Albumin	Transportation and intake of retinol during growth of egg yolk
Ovalbumin Y	Protection of egg yolk from the damage of serpin peptidase
Paraoxonase 2	Catabolism of organophosphorus
Vitellogenin 2	Increased by stimulation of estrogen and decomposed to phosvitin and lipovitellin
Ovocalyxin 32	Resistance to pathogen for influencing survival rate of egg
Ovocleidin 116	Deposition of calcium carbonate in eggshell
Ovocalyxin 36	Related to protection function in eggshell
HEP 21 protein	Related to formation and calcification of eggshell
Fatty acid binding protein 4	Related to deposition and transportation of lipid in muscle cells
Apolipoprotein B	Related to growth and lipid process
Perilipin 1	Related to synthesis and digestion of lipid
Apolipoprotein H	Related to metabolism and transportation of lipid
Hypothetical gene supported by CR353074	Specific expressed in adipose tissue
Secreted phosphoprotein 2	Related to regulation of growth
Proteasome maturation protein	Related to catabolism of protein
Similar to cyclin F	Related to cell cycle transition
Peptidylprolylisomerase D	Related to regulation of apoptosis
Phosphoglycerate kinase 1	Influence of adaptation to hypoxia in chicken embryo
Retinol binding protein 7	Related to energy metabolism and regulation
Sulfotransferase	Related to metabolism of steroid hormone via sulfonation
ATPase, Na+/K+ transporting, alpha 1 polypeptide	Participation in metabolic process, ATPase activity, and cation transportation
Peptidase D	Related to protective immunity effect
PIT 54 protein	Immune function of reducing bacteria and protective effects
CNDP dipeptidase 2	Related to protective immunity effect
Lymphocyte antigen 6 complex	Candidate gene of disease resistance for Marek's disease
Ubiquitin thiolasterase	Related to spermatogenesis and neurogenesis
Signal sequence receptor, beta	A protein expressed in cerebellum
Riboflavin binding protein	Related to regulation of G protein-coupled receptor
Ovoglycoprotein	Responsible to recognition of asymmetry