BMB
reports

Functional analysis of expressed sequence tags from the liver and brain of Korean Jindo dogs

Jae Young Kim[#,1,3], Hye Sun Park[#,1], Dajeong Lim[1], Hong Chul Jang[1], Hae Suk Park[1], Kyung Tai Lee[1], Jong Seok Kim[2], Seok Il Oh[2], Mu Sik Kweon[3], Tae Hun Kim[1] & Bong Hwan Choi[1,]*

[1]Division of Animal Genomics and Bioinformatics, National Institute of Animal Science, Rural Development Administration, Suwon 441-706, [2]Korean Jindo and Domestic Animals Center, Jindo 539-800, [3]Department of Genetic Engineering, Sungkyunkwan University, Suwon 440-746, Korea

We generated 16,993 expressed sequence tags (ESTs) from two libraries containing full-length cDNAs from the brain and liver of the Korean Jindo dog. An additional 365,909 ESTs from other dog breeds were identified from the NCBI dbEST database, and all ESTs were clustered into 28,514 consensus sequences using StackPack. We selected the 7,305 consensus sequences that could be assembled from at least five ESTs and estimated that 12,533 high-quality single nucleotide polymorphisms (SNPs) were present in 97,835 putative SNPs from the 7,305 consensus sequences. We identified 58 Jindo dog-specific SNPs in comparison to other breeds and predicted seven synonymous SNPs and ten non-synonymous SNPs. Using PolyPhen, a program that predicts changes in protein structure and potential effects on protein function caused by amino acid substitutions, three of the non-synonymous SNPs were predicted to result in changes in protein function for proteins expressed by three different genes (TUSC3, ITIH2, and NAT2). [BMB reports 2011; 44(4): 238-243]

INTRODUCTION

The dog was the first domesticated mammal (1) and has had a close relationship with humans for the past 15,000 years. Canines have many domesticated roles, including hunting, working, herding, and companionship (2). Currently, there are more than 300 breeds worldwide, with vast differences in size, shape, and specific behaviors (2). Breeds exhibit many distinct phenotypes, and their traits are largely controlled genetically. In recent years, dogs have become increasingly important as models for human disease due to several advantages. First, dogs generally live in the same environment as their owner, sharing living space and food. Second, dogs and humans share susceptibility to many diseases, including cancers, blindness, cataracts, deafness, epilepsy, heart disease, and genetic disorders and often exhibit symptoms similar to humans (3). Moreover, some diseases can be treated more easily in dogs than in humans (3). Finally, dogs often have a high level of health care, and the canine lifespan is more than five-fold shorter than that of humans, making dogs excellent model organisms.

The complete genome sequence of the domestic dog (Build 2.0) was publicly released in 2005, and was drafted from the DNA sequence of a female Boxer (4). Sequencing was initiated in July 2004 at the Broad Institute of MIT/Harvard and Agencourt Bioscience. This consortium reported an initial compendium of single-nucleotide polymorphisms (SNPs) for the dog population, compiled by comparing draft sequences with sequences from 11 other dog breeds (3). They discovered SNPs by aligning contigs from other species to the Boxer draft sequences (3).

Expressed sequence tags (ESTs) are a rapid and cost-effective method for describing a large number of genetic loci, particularly for comparing species-, tissue-, and organism-specific sequences without fully available sequence information (5). ESTs do not include introns or regions between genes and only include expressed genes.

SNPs present in ESTs have several advantages for use as genetic markers. First, phenotypic frequency calculations are easily performed in any population, as SNPs exhibit di-allelic phenotypic variation (6-9), Second, SNPs are generally more stable markers and are present at higher frequencies than repeat markers (6, 10). Finally, computer analysis systems have been developed to analyze any type of SNP (6, 11). SNP-based analyses have been used to assess many individual genetic characteristics in canines, including coat variation (12), behavior (13-15), and sensitive traits such as smell (16).

The Jindo dog originated on Jindo Island in Korea and is protected as Korea's 53rd natural monument. Korean Jindo dogs are often used for hunting due to their bravery and loyalty. Approximately ten types of Jindo dogs exist, and they can be classified by their coat colors, which are generally yellow or

*Corresponding author. Tel: 82-31-290-1592; Fax: 82-31-290-1592; E-mail: bhchoi@korea.kr
DOI 10.5483/BMBRep.2011.44.4.238
[#]Both authors contributed equally to this work.

Received 23 June 2010 , Accepted 18 January 2011

Keywords: Brain, EST, Korean Jindo dog, Liver, Single Nucleotide Polymorphisms

white but can also be black and feature variations such as a white spot above each eye. The Jindo dog is widely considered an excellent breed; however, few studies have explored its genetics and EST sequences. Therefore, we characterized the ESTs of the Jindo dog using several methods and detected EST-based putative SNPs using PolyBayes. We annotated these potential SNPs by comparing them to NCBI's dbSNP database and the UniProt database. Finally, we identified putative Jindo dog-specific SNPs. Our results provide support for further studies exploring the genetic variation and diversity of dog breeds (2).

RESULTS AND DISCUSSION

EST identification and statistics

We obtained 16,933 high-quality ESTs (mean length, 614 bp) from 20,899 ESTs in two cDNA libraries (brain and liver). We obtained an additional 4,475 clusters from the 16,993 ESTs through clustering analyses using the CAP3 program (17). We generated 1,507 contigs (range, 133-3,427; mean length, 917 bp) consisting of at least two ESTs and 2,968 singletons (range, 101-973; mean length, 571 bp). Coding sequence (CDS) regions were predicted from these clusters by translating all possible reading frames. The CDS regions coded for a mean of 140 amino acids per sequence (5).

Functional annotation and gene identification

Consensus coding sequence (CCDS) regions were identified from the ESTs using BLASTX comparisons between the human and dog. A total of 13,703 human and 13,481 dog were matched CCDSs. The fullness was from the matched CCDS 7,827 (57%) of human and 9,071 (67%) of dog.

The Jindo dog transcripts were aligned against the canine whole genome (Build 2.0) using BLAT (the BLAST-like alignment tool). These results were used to predict the location of genes encoding the Jindo dog transcripts in the whole canine genome. We applied a cutoff of 95% identity and 95% coverage for similar sequences. The most frequently aligned chromosome was chromosome 13, and the most infrequently aligned was chromosome 19. We compared the biasness of brain and liver chromosomes with other dog breeds using the dbEST database (18) (Fig. 1). Fig. 1 shows the distributions of other dog breed ESTs mapped on each of the chromosomes. The red bar shows the total ESTs mapped to each chromosomes, and the blue bar shows those mapped and expressed on each tissue and chromosome.

We performed alignments with mammalian sequences from the Uniprot database using BLASTX (E-value $\leq 1 \times 10^{-5}$) for the 4,475 annotated contigs. This alignment showed contigs with the most assembled ESTs, the identified genes, and their descriptions. The average number of consisted ESTs was nine ESTs and the most frequent number of ESTs in contigs was 2-3 ESTs. The most assembled contig was contig 1369, which consisted of 1,250 ESTs. Contig 1369 was similar to the gene for serum albumin (UniProtId: P49822).

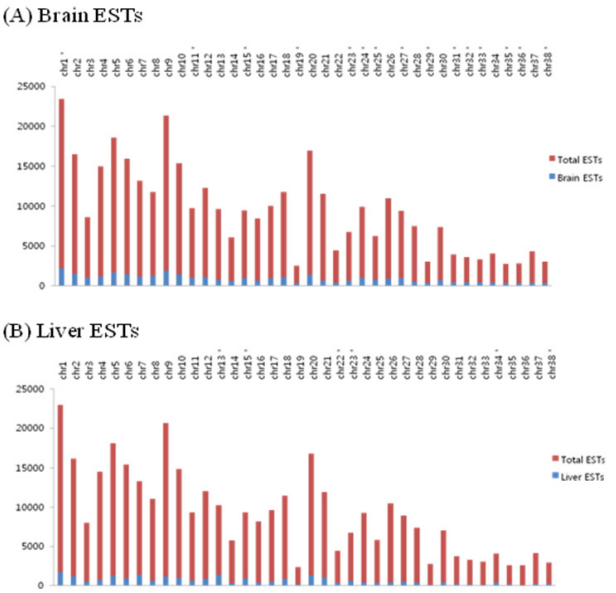

Fig. 1. Chromosome biasness for liver and brain expressed sequence tags (EST). The significant ($P \leq 0.05$) chromosomes are marked with stars.

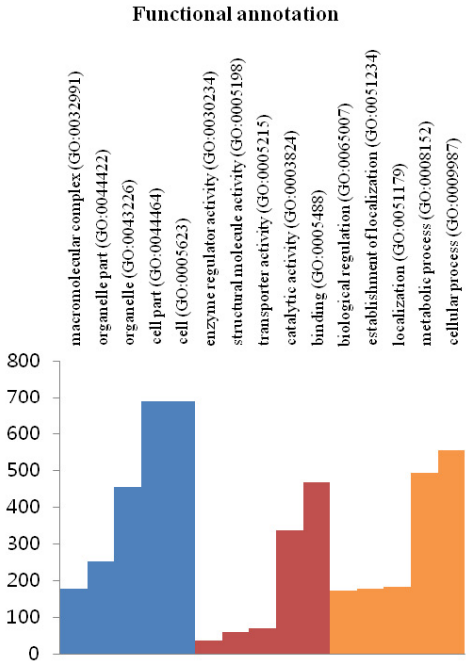

Fig. 2. The most frequently represented gene ontology terms for each of three major gene functions in the contigs. The blue bars are in the cellular component category, red bars are in the molecular function category, and the yellow bars are in the biological processes category.

Table 1. A description of Korean Jindo-dog specific single nucleotide polymorphisms (SNPs)

SNPID	SNP position on chromosome*	Allele	Gene	Gene description
cn2059-1	-	a-	ENSA	Alpha-endosulfine
cn1778-16	62067270(chr4)	t-	Unknown	Unknown
cn1778-19	62067273(chr4)	t-	Unknown	Unknown
cn1328-17	43628801(chr14)	g-	HIBADH	3-hydroxyisobutyrate dehydrogenase, mitochondrial
cn1365-37	-	ct	GDI2	Rab GDP dissociation inhibitor beta
cn479-19	-	t-	MOSC2	MOSC domain-containing protein 2, mitochondrial
cn774-13	38736730(chr21)	tc	Unknown	Unknown
cn703-1	-	at	SH3BGRL3	SH3 domain-binding glutamic acid-rich-like protein 3
cn703-2	-	at	SH3BGRL3	SH3 domain-binding glutamic acid-rich-like protein 3
cn2012-1	58753906(chr18)	ag	LSM8	U6 snRNA-associated Sm-like protein LSm8
cn1811-24	8019570(chr26)	tc	Unknown	Unknown
cn39-23	28672406(chr4)	tc	VDAC2	Voltage-dependent anion-selective channel protein 2
cn3197-3	102473998(chr1)	ga	Chmp2a	Charged multivesicular body protein 2a
cn1957-47	-	g-	Cib2	Calcium and integrin-binding family member 2
cn1957-48	-	gt	Cib2	Calcium and integrin-binding family member 2
cn2186-1	-	ct	SST	Somatostatin
cn3368-6	-	ag	RNASEH2B	Ribonuclease H2 subunit B
cn3338-6	25839687(chr21)	a-	Dgat2	Diacylglycerol O-acyltransferase 2
cn1962-2	54092459(chr11)	ct	NUDT2	Bis(5'-nucleosyl)-tetraphosphatase [asymmetrical]
cn3537-6	54567710(chr3)	gc	MRPS11	28S ribosomal protein S11, mitochondrial
cn3537-11	54576295(chr3)	ga	MRPS11	28S ribosomal protein S11, mitochondrial
cn4834-7	73350068(chr8)	ga	CINP	Cyclin-dependent kinase 2-interacting protein
cn4834-8	73350135(chr8)	ct	CINP	Cyclin-dependent kinase 2-interacting protein
cn315-1	47663004(chr15)	gc	Lsm6	U6 snRNA-associated Sm-like protein LSm6
cn2224-14	38395103(chr18)	ag	RCN1	Reticulocalbin-1
cn1897-5	38802148(chr27)	tc	Gabarapl1	Gamma-aminobutyric acid receptor-associated protein-like 1
cn1876-4	8120244(chr27)	ct	SPATS2	Spermatogenesis-associated serine-rich protein 2
cn460-2	37934073(chr34)	g-	Unknown	Unknown
cn756-22	52217328(chr10)	tc	Unknown	Unknown
cn3238-9	31231490(chr2)	t-	ITIH2	Inter-alpha-trypsin inhibitor heavy chain H2
cn1877-3	55840410(chr18)	tc	Ppp1r14b	Protein phosphatase 1 regulatory subunit 14B
cn1877-4	55841531(chr18)	ga	Ppp1r14b	Protein phosphatase 1 regulatory subunit 14B
cn811-10	28930648(chr3)	gt	Unknown	Unknown
cn2942-2	39813993(chr27)	ct	M6PR	Cation-dependent mannose-6-phosphate receptor
cn2213-23	64621189(chr5)	ac	Unknown	Unknown
cn552-29	-	t-	MRPS18B	28S ribosomal protein S18-2, mitochondrial
cn870-2	57794741(chrX)	ct	GDPD2	Glycerophosphodiester phosphodiesterase domain-containing protein 2
cn238-4	-	a-	Unknown	Unknown
cn238-5	13149633(chr9)	a-	Unknown	Unknown
cn1972-18	88574374(chr1)	gt	Unknown	Unknown
cn5258-12	6313322(chr24)	a-	Nat5	N-acetyltransferase 5
cn890-24	22005785(chr25)	tc	IMP4	U3 small nucleolar ribonucleoprotein protein IMP4
cn653-6	28798214(chr37)	tc	Tuba4a	Tubulin alpha-4A chain
cn3006-31	57305046(chr2)	gt	RAD17	Cell cycle checkpoint protein RAD17
cn3754-3	41836403(chr7)	ga	C1orf55	UPF0667 protein C1orf55
cn1994-7	45649996(chr6)	g-	SARS	Seryl-tRNA synthetase, cytoplasmic
cn2162-8	45078216(chr25)	ga	TRIP12	Probable E3 ubiquitin-protein ligase TRIP12
cn2789-16	55196992(chr18)	-g	ARL2	ADP-ribosylation factor-like protein 2
cn2789-17	55196992(chr18)	-c	ARL2	ADP-ribosylation factor-like protein 2
Cn2789-18	55196992(chr18)	-t	ARL2	ADP-ribosylation factor-like protein 2
Cn2789-19	55196992(chr18)	-g	ARL2	ADP-ribosylation factor-like protein 2
Cn2789-20	55196992(chr18)	-c	ARL2	ADP-ribosylation factor-like protein 2
cn2789-21	55196992(chr18)	-t	ARL2	ADP-ribosylation factor-like protein 2
cn2789-22	55197126(chr18)	ag	ARL2	ADP-ribosylation factor-like protein 2
cn3828-7	17761813(chr38)	tc	EPRS	Bifunctional aminoacyl-tRNA synthetase
cn4569-9	42029403(chr16)	ac	TUSC3	Tumor suppressor candidate 3
cn1786-4	14903754(chr24)	-t	Unknown	Unknown
cn55-16	37062619(chr17)	ag	COX7A2L	Cytochrome c oxidase subunit 7A-related protein, mitochondrial

*SNP position on chromosome: predicted Jindo dog-specific SNP position in the entire canine genome sequence

We also functionally annotated three of the categories on the second level of the gene ontology tree constructed using high similarity and compared them with the UniProt databases. Table S1 presents 34 significant gene ontology terms, including eight in the cellular component category, ten in the molecular function category, and 16 in the biological process category (5). Fig. 2 shows the highest frequency gene ontology terms identified in each category. Cell binding and cellular processes had the largest number of ESTs in terms of the cellular component, molecular function, and biology process categories.

SNP identification and analysis

We predicted 97,835 potential SNPs from the selected consensus sequences. We selected 12,533 high-quality SNPs (with more than four depths and double hits). From these, we estimated that 58 were Jindo dog-specific by comparing them to other breeds. We analyzed these SNPs against the Uniprot database using BLASTX. Table 1 shows data regarding the Jindo dog-specific SNPs, including the predicted SNP position in the genome and allele information.

From the 58 Jindo-specific SNPs, we predicted seven synonymous coding SNPs (cSNPs) and ten non-synonymous cSNPs, which changed the amino acids found in the CDS regions (13). Additionally, nine non-synonymous cSNPs were identified from genes in the UniProt database. Using PolyPhen, a program that predicts changes in protein structure or function caused by amino acid substitutions, three of the non-synonymous SNPs were predicted to potentially affect protein structure or function (cn3238 and cn5258: probably damaging, cn4569: possibly damaging). The remaining SNPs were benign, based on multiple alignments of all SNPs.

Our analysis showed that various effects on protein stability were responsible for the accumulation of slightly deleterious non-synonymous SNPs in Jindo dog-specific genes. In Table 2, we describe three non-synonymous SNPs predicted to have deleterious effects on protein function. These SNPs resided within *ITIH2*, *NAT5*, and *TUSC3*. The inter-α-trypsin inhibitors (ITI) are a family of plasma protease inhibitors that contribute to extracellular matrix stability by covalently linking to hyaluronan. ITI complexes are assembled from a light chain (bikunin, encoded by *AMBP*) and five homologous heavy chains (encoded by *ITIH1, ITIH2, ITIH3, ITIH4,* and *ITIH5*). Thus, ITIH molecules play a particularly important role in in-flammation and carcinogenesis (19). ITIH2 has been identified by biochemical fractionation studies and exists both as a single protein and in a bikunin-bound form. A reduction in ITIH2 expression correlates with brain tumor progression, and targeting factors responsible for ITIH2 loss or exogenous restoration of ITIH2 expression could serve as a potential therapeutic strategy for treating a variety of central nervous system tumors (13).

N-acetyltransferases (NATs) are key enzymes involved in the conjugation of certain drugs and xenobiotics with an arylamine structure (20). NAT5 belongs to the acetyltransferase family, specifically to the ARD1 subfamily. A sequence alignment revealed a high degree of similarity in NAT5 proteins among species, supporting its conserved role as a part of the hARD1-NATH complex. Human NAT5 (hNAT5) is the third component in this complex. Protein acetyltransferases and deacetylases have been implicated in tumorigenesis, apoptosis, and cell cycle regulation.

TUSC3 is a candidate tumor suppressor gene. It is located within a homozygous deleted region in metastatic prostate cancer and is expressed in most non-lymphoid human tissues, including prostate, lung, liver, and colon. *TUSC3* expression has also been detected in many epithelial tumor cell lines. Two transcript variants encoding distinct isoforms have been identified for this gene.

These data could help identify novel or uncharacterized genes for elucidating basic canine genomics by contributing to the number of identified Jindo dog SNPs and ESTs. Since the dog genome was sequenced, an increasing number of studies involving genetic research on dogs have been conducted to understand canine gene function. Our study provides powerful analytical tools for conducting genetic polymorphism analyses in canines. Studies in Jindo dogs and in other domestic animals will expand our basic knowledge of the genetic control

Table 2. PolyPhen predictions for changes in functional effects due to substitutions in Korean Jindo dog specific non-synonymous coding single nucleotide polymorphisms (SNPs)

SNPID	Gene	Position*	Substitution†	Prediction‡
cn1328-17	HIBADH	312	Gly[G]→Asp[D]	Benign
cn2012-1	LSM8	32	Thr[T]→Ala[A]	Benign
cn3368-6	RNASEH2B	266	Lys[K]→Arg[R]	Benign
cn3537-6	MRPS11	52	Glu[E]→Asp[D]	Benign
cn3238-9	ITIH2	94	Asn[N]→Ile[I]	probably damaging
cn5258-12	Nat5	197	Asn[N]→Ile[I]	probably damaging
cn1994-7	SARS	353	Gly[G]→Glu[E]	Benign
cn4569-9	TUSC3	362	Thr[T]→Pro[P]	possibly damaging
cn890-24	IMP4	219	Gln[Q]→Arg[R]	benign

*Position: substitution on the translated consensus sequence position, †Substitution: Change in the amino acid sequence by SNP allele, ‡Prediction: prediction of an effect on protein function or structure by the substitution

of traits. These results can then be used to improve desired traits and to reduce the incidence of disease. The present study provides an essential first step toward establishing an interactive network for gene identification and expression patterns suitable for functional genetics, comparative genomics, and evolutionary analyses of genes and gene families involved in the developmental and degenerative processes of the brain and liver. With the growth of public SNP data and improvements in the quality of SNP databases, functional genetics using SNPs could play an important role in furthering our understanding of the inheritance and composition of canine phenotypes.

MATERIALS AND METHODS

Sample preparation and total RNA isolation

Liver and brain tissue of three Korean Jindo dogs was excised immediately after euthanasia, snap frozen in liquid nitrogen, and stored at −70°C until RNA extraction. Total RNA was extracted from all samples using an RNeasy Lipid Tissue Midi kit (Qiagen, Valencia, CA, USA), according to the manufacturer's instructions.

Full-length cDNA library construction and cDNA sequencing

We constructed full-length enriched cDNA libraries from a large amount of mRNA (21-23). The ligated cDNA was then transformed into *E. coli* Top 10F′ (Invitrogen, Carlsbad, CA, USA) by electroporation (Gene Pulser II BioRad, Hercules, CA, USA). The cDNA insert lengths were investigated for about 21,000 randomly selected clones. The extracted plasmids were digested with EcoRI and NotI restriction enzymes. The obtained plasmid DNA products were sequenced using the ABI Prism Big Dye Terminator Cycle Sequencing kit version 3.1 (Applied Biosystems, Foster City, CA, USA) in an ABI 3730XL sequencer (Applied Biosystems). Reaction conditions were as follows: 2.5 μl canine plasmid DNA as the template, 0.875 μl 5× buffer containing 0.5 μl T7 Universal Primer (0.6 pmol/μl), 0.25 μl Big dye (ABI v3.1), and water to a total volume of 5 μl with the following parameters: 95°C for 10 seconds, 50°C for 5 seconds, and 60°C for 4 minutes.

Data preparation and preprocessing

We downloaded the entire canine genome sequence databases from UCSC and NCBI (Build 2.0). We also downloaded canine ESTs and SNPs from the dbEST (accession: U39396-AJ407864) database and the dbSNP (Build 131) database at NCBI. We generated 20,899 ESTs from Jindo-dog brain and liver cDNA libraries and performed EST pre-processing to remove redundancy data and to obtain high-quality sequences. The ESTs were base-called from chromat files using PHRED (24) and then were trimmed over than 0.05 of base call error probability rate. The sequences were clipped for vector sequences using cross-match and were masked using Repeat Masker (25). We used the default options in this program. The parameters for the Cross match program were "cross_match -canine_ests.raw −pCNS-D2 −minmatch 12 −minscore 20 −screen > screen.out" and "Repeat maskers was RepeatMasker −q −x −species dog canine_ests.seq". The registered accession numbers from the NCBI dbEST database were GR885592-GR902300.

Gene ontology annotation and functional analysis

To determine the position of our transcripts on canine chromosomes, we aligned the canine ESTs, contigs, and singletons using BLAT (26) against the entire canine genome (Build 2.0) in the UCSC database. To annotate our transcripts with Gene Ontology, we aligned our transcripts against 495,880 mammalian protein sequences from UniProt (release 15) database using BLASTX (27) with a cutoff of 1×10^{-5}. These identified ESTs were categorized to the second level in each of three categories (biological processes, molecular functions, cellular components) using the database at the Gene Ontology Consortium website (http://geneontology.org/). Fisher's exact test was used to test if function bias existed in the ESTs. As there was a multiple test problem, the false discovery rate correction was applied for the test. Results of multiple tests can result in an unexpected number of type I errors.

SNP detection and possible amino acid changes

To identify SNPs from the EST database, we clustered the ESTs and added other dog-breed EST data from the dbEST database using the StackPack program (28). We selected 7,305 consensus sequences that consisted of at least five ESTs from 28,514 consensus sequences. We detected SNPs from the used the consensus sequences using PolyBayes program. The PolyBayes is an anchored multiple sequences alignment program (29). We determined high-quality SNPs with at least a depth of four. We detected Jindo-dog specific SNPs against ESTs of other breeds from these high-quality SNPs. Flanking SNP sequences were mapped against the canine genome sequences using the BLAT program to determine the genomic position of the putative SNPs. We defined the CDS region from a six frame translation (30) and the cSNPs. We further analyzed the SNPs showing only Jindo-dog ESTs and defined synonymous and non synonymous SNPs (31). The damaging effect of amino acid substitutions were predicted from SNPs using PolyPhen (13, 30). PolyPhen predicts functional effects or structures using non synonymous SNPs (13).

Acknowledgments

This work was supported by the "High-throughput Sequencing and Variation for Genetic Markers Development in Korean Native Animals" project of the National Institute of Animal Science, RDA, Korea.

REFERENCES

1. Cruz, F., Vila, C. and Webster, M. T. (2008) The legacy of

domestication: accumulation of deleterious mutations in the dog genome. *Mol. Biol. Evol.* **25**, 2331-2336.
2. Palmer, L. E., O'Shaughnessy, A. L., Preston, R. R., Santos, L., Balija, V. S., Nascimento, L. U., Zutavern, T. L., Henthorn, P. S., Hannon, G. J. and McCombie, W. R. (2003) A survey of canine expressed sequence tags and a display of their annotations through a flexible web-based interface. *J. Hered.* **94**, 15-22.
3. Lindblad-Toh, K., et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803-819.
4. Ostrander, E. A. and Wayne, R. K. (2005) The canine genome. *Genome Res.* **15**, 1706-1716.
5. Frentiu, F. D., Adamski, M., McGraw, E. A., Blows, M. W. and Chenoweth, S. F. (2009) An expressed sequence tag (EST) library for Drosophila serrata, a model system for sexual selection and climatic adaptation studies. *BMC Genomics* **10**, 40.
6. Picoult-Newberg, L., Ideker, T. E., Pohl, M. G., Taylor, S. L., Donaldson, M. A., Nickerson, D. A. and Boyce-Jacino, M. (1999) Mining SNPs from EST databases. *Genome Res.* **9**, 167-174.
7. Barbazuk, W. B., Emrich, S. J., Chen, H. D., Li, L. and Schnable, P. S. (2007) SNP discovery via 454 transcriptome sequencing. *Plant J.* **51**, 910-918.
8. Hayes, B. J., Nilsen, K., Berg, P. R., Grindflek, E. and Lien, S. (2007) SNP detection exploiting multiple sources of redundancy in large EST collections improves validation rates. *Bioinformatics* **23**, 1692-1693.
9. Kwok, P. Y., Carlson, C., Yager, T. D., Ankener, W. and Nickerson, D. A. (1994) Comparative analysis of human DNA variations by fluorescence-based sequencing of PCR products. *Genomics* **23**, 138-144.
10. Wang, D. G., Fan, J. B., Siao, C. J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M. S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T. J., Lipshutz, R., Chee, M. and Lander, E. S. (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**, 1077-1082.
11. Hacia, J. G., Brody, L. C., Chee, M. S., Fodor, S. P. and Collins, F. S. (1996) Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. *Nat. Genet.* **14**, 441-447.
12. Cadieu, E., Neff, M. W., Quignon, P., Walsh, K., Chase, K., Parker, H. G., Vonholdt, B. M., Rhue, A., Boyko, A., Byers, A., Wong, A., Mosher, D. S., Elkahloun, A. G., Spady, T. C., Andre, C., Lark, K. G., Cargill, M., Bustamante, C. D., Wayne, R. K. and Ostrander, E. A. (2009) Coat variation in the domestic dog is governed by variants in three genes. *Science* **326**, 150-153.
13. Vage, J. and Lingaas, F. (2008) Single nucleotide polymorphisms (SNPs) in coding regions of canine dopamine- and serotonin-related genes. *BMC Genet.* **9**, 10.
14. Spady, T. C. and Ostrander, E. A. (2008) Canine behavioral genetics: pointing out the phenotypes and herding up the genes. *Am. J. Hum. Genet.* **82**, 10-18.
15. Arata, S., Ogata, N., Shimozuru, M., Takeuchi, Y. and Mori, Y. (2008) Sequences and polymorphisms of the canine monoamine transporter genes SLC6A2, SLC6A3, and SLC6A4 among five dog breeds. *J. Vet. Med. Sci.* **70**, 971-975.
16. Lesniak, A., Walczak, M., Jezierski, T., Sacharczuk, M., Gawkowski, M. and Jaszczak, K. (2008) Canine olfactory receptor gene polymorphism and its relation to odor detection performance by sniffer dogs. *J. Hered.* **99**, 518-527.
17. Huang, X. and Madan, A. (1999) CAP3: a DNA sequence assembly program. *Genome Res.* **9**, 868-877.
18. Nakaya, H. I., Amaral, P. P., Louro, R., Lopes, A., Fachel, A. A., Moreira, Y. B., El-Jundi, T. A., da Silva, A. M., Reis, E. M. and Verjovski-Almeida, S. (2007) Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. *Genome Biol.* **8**, R43.
19. Werbowetski-Ogilvie, T. E., Agar, N. Y., Waldkircher de Oliveira, R. M., Faury, D., Antel, J. P., Jabado, N. and Del Maestro, R. F. (2006) Isolation of a natural inhibitor of human malignant glial cell invasion: inter alpha-trypsin inhibitor heavy chain 2. *Cancer Res.* **66**, 1464-1472.
20. Walker, K., Ginsberg, G., Hattis, D., Johns, D. O., Guyton, K. Z. and Sonawane, B. (2009) Genetic polymorphism in N-Acetyltransferase (NAT): Population distribution of NAT1 and NAT2 activity. *J. Toxicol. Environ. Health B. Crit. Rev.* **12**, 440-472.
21. Oh, J. H., Kim, Y. S. and Kim, N. S. (2003) An improved method for constructing a full-length enriched cDNA library using small amounts of total RNA as a starting material. *Exp. Mol. Med.* **35**, 586-590.
22. Sasako, M., Sano, T., Katai, H., Okajima, K. and Maruyama, K. (1994) Overview of clinical trials on adjuvant chemotherapy for curatively resected gastric cancer. *Gan. To. Kagaku. Ryoho.* **21(Suppl 3)**, 384-394.
23. Dirisala, V. R., Kim, J., Park, K., Kim, N., Lee, K. T., Oh, S. J., Oh, J. H., Kim, N. S., Um, S. J., Lee, H. T., Kim, K. I. and Park, C. (2005) cSNP mining from full-length enriched cDNA libraries of the Korean native pig. *Korean J. Genet.* **27**, 329-335.
24. Ewing, B., Hillier, L., Wendl, M. C. and Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175-185.
25. Jurka, J. (2000) Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet.* **16**, 418-420.
26. Kent, W. J. (2002) BLAT–the BLAST-like alignment tool. *Genome Res.* **12**, 656-664.
27. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.
28. Christoffels, A., van Gelder, A., Greyling, G., Miller, R., Hide, T. and Hide, W. (2001) STACK: Sequence Tag Alignment and Consensus Knowledgebase. *Nucleic Acids Res.* **29**, 234-238.
29. Useche, F. J., Gao, G., Harafey, M. and Rafalski, A. (2001) High-throughput identification, database storage and analysis of SNPs in EST sequences. *Genome Inform.* **12**, 194-203.
30. Ramensky, V., Bork, P. and Sunyaev, S. (2002) Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.* **30**, 3894-3900.
31. Sunyaev, S., Ramensky, V., Koch, I., Lathe, W., Kondrashov, A. S. and Bork, P. (2001) Prediction of deleterious human alleles. *Hum. Mol. Genet.* **10**, 591-597.