REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	6.25.03	Final report 6.1.01-3.31.03

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
Radiation Hybrid Mapping of Canine BACs	DAAD 19-01-0658
6. AUTHOR(S) Elaine A. Ostrander, Ph.D.	

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Fred Hutchinson Cancer Research Center 1100 Fairview Ave. N., D4-100, Seattle, WA 98109	6249

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING / MONITORING AGENCY REPORT NUMBER
U. S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	42431.4-LS

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT	12 b. DISTRIBUTION CODE
Approved for public release; distribution unlimited.	

13. ABSTRACT (Maximum 200 words)

Our proposal focused on the placement of several hundred Bacterial Artificial Chromosome (BAC) clones on the canine genome map, with the stated goal of developing resources needed for positional cloning of canine disease genes and genes important in mammalian morphology and development. This resource will prove invaluable in the mapping genes that are of interest to human health and biology including those associated with phenotypic variation. Towards this end we have now produced sequence from one or both ends of 4374 BACs. We have placed a total of 1910 of these BACs on the canine genome radiation hybrid map. BAC ends with microsatellite genetic markers were given high priority for mapping as they provide an additional resource for genome wide scans on pedigrees of interest. BACs are well distributed throughout all chromosomes. Through publication, web sites, and talks at high profile meetings we have distributed this information to colleagues. Finally, as part of our effort to make the map most useful to investigators in the field, we recently conducted a detailed analysis of human chromosome 1p, integrating a large number of canine specific genes as well as markers and BAC ends.

14. SUBJECT TERMS	15. NUMBER OF PAGES
BAC-ends, canine, genome map, phenotypic variation, radiation hybrid	11
	16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION ON THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

June 24, 2003 Final Progress Report

US Army Grant DAAD19-01-0658

Co-Investigators: Elaine A. Ostrander, Ph.D and Francis Galibert, Ph.D.

Radiation Hybrid Mapping of Canine BACs

Prepared by:

Principal Investigator

Elaine A. Ostrander Ph.D.
Member, Division of Human Biology and Clinical Research
Fred Hutchinson Cancer Research Center
1100 Fairview Ave. N., D4-100
Seattle WA 98109
Phone: 206.667.6979; FAX 206.667.6396
eostrander@fhcrc.org

Section (2) Table of Contents

Section (3) List of Appendices

A. Guyon R., Lorentzen T.D., Hitte C., Kim L., Cadieu E., Parker H.G., Quignon P., Lowe J.K., Renier C., Gelfenbeyn B., Vignaux G., DeFrance H.B., Gloux S., Mahairas G.G., André C., Galibert F., Ostrander E.A. (2003). A 1 Mb resolution radiation hybrid map of the canine genome. Proc. Natl. Acad. Sci. U.S.A. *100*: 5296-5301.

B. Hitte C., Lorentzen T., Guyon R., Kim L., Cadieu E., Parker H.G., Quignon P., Lowe J.K., Gelfenbeyn B., Andre C., Ostrander E.A., Galibert F. (2003). Comparison of the MultiMap and TSP/CONCORDE packages for constructing radiation hybrid maps. J Hered. *94*:9-13.

C. Guyon, R., Kirkness, E.F., Lorentzen, T.D., Hitte, C., Comstock, K.E., Quignon, P., Derrien, T., André, C., Fraser, C.M., Galibert, F., Ostrander E.A. Comparative Mapping of Human Chromosome 1p and the Canine Genome. Submitted to Genome Research

D. Lingaas, F., Comstock, K.E., Kirkness, E.F., Sørensen, A., Aarskaug, T., Hitte, C., Nickerson, M.L., Moe, L., Schmidt, L.S., Thomas, R., Breen, M., Galibert, F., Zbar, B., Ostrander, E.A. A Mutation in the Canine BHD Gene is Associated with Hereditary Multifocal Renal Cystadenocarcinoma and Nodular Dermatofibrosis in the German Shepherd Dog. Submitted to Human Molecular Genetics

Section (4) Statement of Problem Studied

In the following document we summarize progress on our goal of constructing a Bacterial Artificial Chromosome (BAC) map of the canine genome. This will allow investigators interested in identifying genes important in human and canine development, health and biology the means to move from linked marker to gene by establishing the relationship of the canine map to the more complete human and mouse physical maps. This, in turn, will facilitate development of contigs across regions of linkage and subsequent selection of expressed sequence tags (ESTs), cDNAs, and candidate genes.

Our initial aims, requested in a three year grant, were to:

Aim I) End-sequence at least 2,250 canine Bacterial Artificial Chromosome Clones (BAC-ends) over the three years of this grant.

Aim II) Map a minimum of 1,600 BACs-ends using an existing 5,000-rad radiation hybrid (RH) panel over the three years of this grant.

Aim III) Integrate the new BAC-end map data with the existing canine map and distribute the above information to the research community via manuscripts, presentations, and web sites.

An award of 18 rather then 36 months forced us to revise out aims with a goal of placing approximately 1300-1500 BACs on the canine map. We are pleased to report that this goal has more then been achieved.

Section (5) Summary of Most Important Results

To best conclude our work, we have summarized it by initial aim. Our progress to date has been substantial and is enumerated as follows.

Aim 1- End sequencing of BACs.

We based our planned aims on BACs because extensive research has shown they are ideal for mapping experiments because they can carry relatively large inserts, on the order of 180 kb, and because of their low copy number per cell they do not tend to recombine to produce deleted or chimeric clones (1). Indeed, BACs have become the mainstay of the human, *Arabidopsis*, plant, fly and mouse genome projects (2) (3) (4).

In 1999, the Ostrander lab was part of a collaboration, led by Dr. Emmanuel Mignot at Stanford University, to construct and distribute a dense, high quality canine BAC library (5). The library and filters were made under the direction of Dr. Peter DeJong at Roswell Park. Approximately 165,888 canine BAC clones are contained in the library, and they are arrayed in 432 384-well microtiter dishes. The library is gridded onto nine high-density hybridization filters with each clone in a prearranged duplicate pattern. Analysis of randomly selected clones indicates a mean insert size of 155 kb and predicts an 8.1-fold coverage of the canine genome. No chimerism was detected in FISH studies of 60 BAC clones (5). The gridded library is in place in the Ostrander lab as well as in several other labs in the U.S. and Europe.

When we began this project we and our collaborators first attempted sequencing of an initial set of 2016 BACs. After automated preparation and standard sequencing, trace files representing each BAC-end sequence were imported from ABI sequencers. Each sequence was examined for homology to cloning vectors, *E. coli* and repetitive DNA sequence. We obtained high quality sequence from either one or both ends of 1504 BACs (766 for one end only, 738 for both ends). Average read lengths were in excess of 700 bp. The 4032 sequences generated had an average of 342 bases with Phred scores ≥20. For BAC-end mapping, high quality (HQ) sequence is defined as having 100 continuous sequences with Phred scores of 20 or greater. The base calling program Phred computes the probability of an error in the base call at each position, and converts this to a quality value (6, 7). A Phred quality score of 20 or higher indicates that there is less than a 1:100 chance that the base-call is incorrect. PCR primers were then selected for genotyping using standard selection programs, i.e. Oligo3 or Primer3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi).

Following this, we developed a collaboration with Ewen Kirkness at The Institute for Genomic Research (TIGR) to generate sequence for another set of several thousand BACs. This was successful; we achieved high quality sequence from both ends of 2281 new BACs, and high quality sequence from only one end of 589 additional BACs. In total, therefore we have produced sequence from one or both ends of 4374 BACs. We have thus more then achieved Aim 1.

Aim 2-Map a minimum of 1,600 BACs-ends using an existing 5,000-rad radiation hybrid (RH) panel.

All mapping experiments were done on a whole-genome radiation hybrid (WGRH) panel created by fusing canine fibroblasts, which had been gamma irradiated at 5000 rads, to a recipient thymidine kinase deficient hamster cell line (HTK3-1) isolated and cloned from A2H (Chinese Hamster ovary cells) (8). The resulting panel of 118 hybrid lines has been used by

nearly all investigators in the field interested in building maps of localized regions as well as by us for all versions of the canine genome map (9) (10) (11).

Relatively standard genotyping protocols were employed in these experiments. After masking for repetitive sequences, primers for RH mapping were selected from each BAC-end using standard selection programs, i.e. Oligo3 or Primer3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). All selected primers will preferentially have a size of 23 bp in order to minimize problems associated with non-specific amplification, thus optimizing the probability that unique size products will be produced with dog DNA in the presence of hamster DNA. Once primers are selected, genotyping and mapping are done using existing infrastructure that was originally developed and optimized by the Galibert lab (9) (12) and which is currently in place in both Ostrander and Galibert labs (10) (11) (13). Primers defining each BAC-end were selected from sequence with the highest number of high quality (HQ) bases. HQ sequence was defined as having 100 continuous sequences with Phred scores of 20 or greater. Only one set of primers was used to genotype each BAC; primers were designed from the opposite end of the insert were used for genotyping only if the first pair yielded poor quality data. For approximately 65% of BACs genotyping was performed in duplicate to ensure high quality.

Novel markers were incorporated into the previous 1500 marker RH data set (11) by pairwise calculations using MultiMap software (14) at a Lod threshold $\geq$ 8.0. A total of 3162 markers could be clustered into RH groups. RH groups were ordered using the Traveling Salesman Problem (TSP) approach as specified by the CONCORDE computer package (15). TSP/CONCORDE computes five independent RH maps; three are variants of the maximum likelihood estimate approach (MLE) and two are constructed using obligate chromosome breaks (OCB). The resulting maps were evaluated to produce a consensus map using a novel method developed by us as part of the US Army funded BAC mapping effort (16). A copy of this paper is included; please note appropriate acknowledgements. For markers whose map position was not well supported, genotyping data was re-examined and genotypes repeated. When no erroneous genotypes were observed, the problematic linkage group was split into two or more RH groups using the MultiMap algorithm and a Lod threshold of >9.0 (14).

Inter-marker distances were determined with the rh_tsp_map1.0 version of TSP/CONCORDE which delivers map positions in arbitrary units. For each chromosome the sum of the arbitrary units was converted into kb using the known physical size of each chromosome, as determined by cytofluorimetry (17). When more than one RH group was assigned to a chromosome, 350 Units were added for each gap, corresponding to the upper limit of our ability to detect linkage between adjacent markers. As of June 2003, we have mapped a total of 1910 BACs on the canine RHDF 5000 rad panel, again surpassing our projected goals.

Aim III- Integrate the new BAC-end map data with the existing canine map and distribute the above information to the research community via manuscripts, presentations, and web sites.

Towards this end, we recently presented a comprehensive radiation hybrid (RH) map of the canine genome composed of 3270 markers including 1596 microsatellite-based markers, 106 Sequence-Tagged Sites (STS), 900 cloned gene sequences and Expressed Sequence Tags (ESTs), 668 canine-BAC ends from the above described work (13). The work was published in the *Proceedings Of the National Academy of Sciences*, and was contributed by Nobel Prize winner E. Donnell Thomas, M.D., in recognition of the significant achievement the paper represented.

In assembling the 3200 marker map, inclusive of the above mentioned BACs, pairwise linkage analysis at a Lod threshold ≥8.0 using MultiMap allowed the localization of 3,162 markers to the 38 autosomes and sex chromosomes, leaving only 17 orphan RH groups and 108 unlinked markers. Of the 17 orphan groups, comprising 2 to 19 markers, 14 could be incorporated into RH groups already assigned to chromosomes using two-point analyses with Lod scores between 5.0 and 8.0. For eight groups the resulting map position is in full agreement with predictions from syntenic human data and for one group a synteny break is introduced. The three remaining orphan RH groups contain only 14 markers. Thus the vast majority of markers are now assigned on the map.

The number of markers assigned to each autosome ranged from 156 markers at 147 unique positions on chromosome 1 (CFA 1) to a minimum of 25 markers at 24 positions (CFA 38). The smallest canine chromosome, the Y, has 18 markers (Table 1). The total map size for individual autosomes ranges from 12,353 U (CFA 1) to 1783 U (CFA 38 The total size of the complete RH map is 227,477 U. The 3270 markers map to 3009 unique positions; 261 markers (8%) are co-positioned. In one case, CFA 35, five independent markers co-localize to a unique position. The average inter-marker distance of the map is 78 U, or approximately 900 kb. The present map, therefore, represents a global two-fold increase in marker density compared to previous iterations of the map (11), with a concomitant 1.5-fold increase in the number of microsatellite markers, a 2.8-fold increase in EST/gene markers and a novel set of mapped BAC-end sequences.

We note specifically that BAC-ends are randomly distributed throughout all chromosomes; ranging from one on CFA Y to 42 on CFA 1. These 668 mapped BAC-ends constitute an initial framework of clones for anchoring the canine physical map and provide a format for positional cloning studies. A subset of 39 mapped BAC clones also contained microsatellites within the end sequences making them particularly useful to mappers and those engaged in positional cloning experiments. These are indicated in the manuscript of Guyon et al (2003) and all associated web figures with a star (18). We have included a copy of the relevant manuscript, noting the US Army acknowledgements, in this packet. All data generated to date, including the complete canine maps can be viewed on our websites: (http://www-recomgen.univ-rennes1.fr/doggy.html) and (http://www.fhcrc.org/science.dog_genome/dog.html.). We note also that this work is being featured at the recent Cold Spring Harbor 68th Symposium on the Genome of *Homo sapiens*. Drs. Ostrander and Galibert have been awarded a plenary talk on May 30, 2003 to summarize the status of the canine map, inclusive of the US Army funded BAC mapping effort.

We have thus, through publication, web sites, and talks at high profile meetings achieved the goals of Aim 3. One additional publication is currently in preparation with Dr. Kirkness that will include data on the remaining 1242 BACs that are currently mapped but were not included in the initial 3200 marker map paper (Kirkness, Ostrander and Galibert Laboratories, In Preparation). These data were largely generated during the assembly of that paper and could not be included at the time.

Additional work on Human Chromosome 1p

As part of our effort to make the map most useful to investigators in the field, we recently conducted a detailed analysis of human chromosome 1p, integrating a large number of canine specific genes as well as markers and BAC ends. Specifically, we defined the evolutionary relationships between the canine genome and human chromosome 1p (HSA1p). The definition and mapping of 120 novel canine genes, orthologous to HSA1p genes, allowed identification of

seven conserved segments within five chromosomal regions (Canis familiaris chromosomes (CFA) 2, 5, 6, 15 and 17) (18). This paper has been submitted for publication, and is also appended. Note again the appropriate acknowledgement of the US Army Grant.

Summary:

Our proposal focused on the placement of several hundred BAC clones on the canine genome map, with an eye towards developing resources needed for positional cloning of canine disease genes. As we summarized above, this resource will prove invaluable in the mapping of canine disease genes that are of interest to human health and biology. But perhaps of more interest to the U.S. Army, is the utility of the resource for mapping traits associated with phenotypic variation, such as leg length, body shape and size, and behavior. Overall such genes could obviously play a key role in defining physical and mental attributes that would affect ones ability to successfully complete physical tasks, an issue of interest to the army. Thus, the work completed here, we believe, will further our understanding of the genetics of both human and canine health and biology.

(6) Listing of All Publications

(a) Peer Reviewed Journals

Guyon R., Lorentzen T.D., Hitte C., Kim L., Cadieu E., Parker H.G., Quignon P., Lowe J.K., Renier C., Gelfenbeyn B., Vignaux G., DeFrance H.B., Gloux S., Mahairas G.G., André C., Galibert F., Ostrander E.A. (2003). A 1 Mb resolution radiation hybrid map of the canine genome. Proc. Natl. Acad. Sci. U.S.A. *100*: 5296-5301.

Hitte C., Lorentzen T., Guyon R., Kim L., Cadieu E., Parker H.G., Quignon P., Lowe J.K., Gelfenbeyn B., Andre C., Ostrander E.A., Galibert F. (2003). Comparison of the MultiMap and TSP/CONCORDE packages for constructing radiation hybrid maps. J Hered. *94*:9-13.

(b) Conference Proceedings

Guyon R., Kirkness E. F., Lorentzen T. D., Hitte C., Comstock K. E., Quignon P., Derrien T., Andréa C., Fraser C. M., Galibert F., and Ostrander E. A. Comparative mapping of human chromosome 1p and the canine genome. *Cold Spring Harbor 68th Symposium on the Genome of Homo Sapiens*, May 2003. Cold Spring Harbor, NY

Guyon R., Lorentzen T.D., Hitte C., Kim L., Cadieu E., Parker H.G., Quignon P., Lowe J.K., Renier C., Gelfenbeyn B., Vignaux G., DeFrance H.B., Gloux S., Mahairas G.G., André C., Galibert F., Ostrander E.A. A 1 Mb Resolution Map of the Canine Genome. *Cold Spring Harbor Genome Mapping and Sequencing Meeting*, May 2002. Cold Spring Harbor, NY

(c) Presentations at Meetings but not Published

Ostrander E.A, Guyon R, Kirkness EF, Hitte C, Comstock K, Cadieu E, Parker H G, Quignon P, Lorentzen T, Renier C, Lowe JK, Gloux S , Vignaux F., Kim L, André C, Galibert F. High Resolution RH Mapping of the Dog Genome and Its Application to the Positional Cloning of Cancer Genes. *The Canine and Feline Genomes.* May 2002. St. Louis MO.

(d) Submitted but Not Published

Guyon R., Kirkness E. F., Lorentzen T. D., Hitte C., Comstock K. E., Quignon P., Derrien T., André C., Fraser C. M., Galibert F., and Ostrander E. A. Comparative mapping of human chromosome 1p and the canine genome (Submitted).

(e) Technical Reports

None

(7) Listing Of Participating Scientific Personnel Showing Advanced Degrees Earned by them while Employed on Project

No staff earned degrees while on project

(8) Report of Inventions

None

(9) Bibliography

1. Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector, Proc Natl Acad Sci U S A. *89:* 8794-7., 1992.
2. Kim, U. J., Shizuya, H., Kang, H. L., Choi, S. S., Garrett, C. L., Smink, L. J., Birren, B. W., Korenberg, J. R., Dunham, I., and Simon, M. I. A bacterial artificial chromosome-based framework contig map of human chromosome 22q, Proc Natl Acad Sci U S A. *93:* 6297-6301, 1996.
3. Korenberg, J. R., Chen, X. N., Sun, Z., Shi, Z. Y., Ma, S., Vataru, E., Yimlamai, D., Weissenbach, J. S., Shizuya, H., Simon, M. I., Gerety, S. S., Nguyen, H., Zemsteva, I. S., Hui, L., Silva, J., Wu, X., Birren, B. W., and Hudson, T. J. Human genome anatomy: BACs integrating the genetic and cytogenetic maps for bridging genome and biomedicine, Genome Res. *9:* 994-1001, 1999.

4. Budiman, M., Mao, L., Wood, T. C., and Wing, R. A. A deep-coverage tomato BAC library and prospects toward development of an STC framework for genome sequencing., Gen Res. *10:* 129-136, 2000.
5. Li, R., Mignot, E., Faraco, J., Kadotani, H., Cantanese, J., Zhao, B., Lin, X., Hinton, L., Ostrander, E. A., Patterson, D. F., and de Jong, P. J. Construction and characterization of an eightfold redundant dog genomic bacterial artificial chromosome library, Genomics. *58:* 9-17, 1999.
6. Ewing, B., Hillier, L., Wendl, M. C., and Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment, Genome Res. *8:* 175-85., 1998.
7. Ewing, B. and Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities, Genome Res. *8:* 186-94., 1998.
8. Vignaux, F., Hitte, C., Priat, C., Chuat, J. C., Andre, C., and Galibert, F. Construction and optimization of a dog whole-genome radiation hybrid panel, Mamm Genome. *10:* 888-894, 1999.
9. Priat, C., Hitte, C., Vignaux, F., Renier, C., Jiang, Z., Jouquand, S., Cheron, A., Andre, C., and Galibert, F. A whole-genome radiation hybrid map of the dog genome., Genomics. *54:* 361-378, 1998.
10. Mellersh, C. S., Hitte, C., Richman, M., Vignaux, F., Priat, C., Jouquand, S., Werner, P., André, C., DeRose, S., Patterson, D. F., Ostrander, E. A., and Galibert, F. An integrated linkage-radiation hybrid map of the canine genome, Mammalian Genome. *11:* 120-130, 2000.
11. Breen, M., Jouquand, S., Renier, C., Mellersh, C. S., Hitte, C., Holmes, N. G., Cheron, A., Suter, N., Vignaux, F., Bristow, A. E., Priat, C., McCann, E., Andre, C., Boundy, S., Gitsham, P., Thomas, R., Bridge, W. L., Spriggs, H. F., Ryder, E. J., Curson, A., Sampson, J., Ostrander, E. A., Binns, M. M., and Galibert, F. Chromosome-specific single-locus FISH probes allow anchorage of an 1800-marker integrated radiation-hybrid/linkage map of the domestic dog genome to all chromosomes, Genome Res. *11:* 1784-95., 2001.
12. Vignaux, F., Priat, C., Jouquand, S., Hitte, C., Jiang, Z., Cheron, A., Renier, C., Andre, C., and Galibert, F. Toward a dog radiation hybrid map., Journal of Hereditary. *90:* 62-7, 1999.
13. Guyon, R., Lorentzen, T. D., Hitte, C., Kim, L., Cadieu, E., Parker, H. G., Quignon, P., Lowe, J. K., Renier, C., Gelfenbeyn, B., Vignaux, G., DeFrance, H. B., Gloux, S., Mahairas, G. G., André, C., Galibert, F., and E.A., O. A 1 Mb resolution radiation hybrid map of the canine genome., Proc. Natl. Acad. Sci. U.S.A. In Press, 2003.
14. Matise, T. C., Perlin, M., and Chakravarti, A. Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linkage map, Nature Genet. *6:* 384-390, 1994.
15. Agarwala, R., Applegate, D. L., Maglott, D., Schuler, G. D., and Schaffer, A. A. A fast and scalable radiation hybrid map construction and integration strategy, Genome Res. *10:* 350-64., 2000.
16. Hitte, C., Lorentzen, T., Guyon, R., Kim, L., Cadieu, E., Parker, H., Quignon, P., Lowe, J., Gelfenbeyn, B., Andre, C., Ostrander, E., and Galibert, F. Comparison of the MultiMap and TSP/CONCORDE packages for constructing radiation hybrid maps., J. Heredity, 2003.

17. Langford, C. F., Fischer, P. E., Binns, M. M., Holmes, N. G., and Carter, N. P. Chromosome-specific paints from a high-resolution flow karyotype of the dog, Chromosome Res. *4:* 115-123, 1996.
18. Guyon, R., Kirkness, E. F., Lorentzen, T. D., Hitte, C., Comstock, K. E., Quignon, P., Derrien, T., Andréa, C., Fraser, C. M., Galibert, F., and Ostrander, E. A. Comparative mapping of human chromosome 1p and the canine genome. Submitted.

Appendix A
Elaine A. Ostrander, Ph.D.
DAAD-19-01-0658

A 1-Mb resolution radiation hybrid map of the canine genome

Richard Guyon*, Travis D. Lorentzen†, Christophe Hitte*, Lisa Kim†, Edouard Cadieu*, Heidi G. Parker†, Pascale Quignon*, Jennifer K. Lowe†, Corinne Renier*‡, Boris Gelfenbeyn†, Françoise Vignaux*, Hawkins B. DeFrance†, Stephanie Gloux*, Gregory G. Mahairas§, Catherine André*, Francis Galibert*¶, and Elaine A. Ostrander†¶

*Unité Mixte de Recherche 6061, Centre National de la Recherche Scientifique, Génétique et Développement, Faculté de Médecine, 35043 Rennes Cédex, France; and †Clinical and Human Biology Divisions, Fred Hutchinson Cancer Research Center, P.O. Box 19024, D4-100, Seattle, WA 98109-1024

Communicated by E. Donnall Thomas, Fred Hutchinson Cancer Research Center, Seattle, WA, February 19, 2003 (received for review December 17, 2002)

The purebred dog population consists of >300 partially inbred genetic isolates or breeds. Restriction of gene flow between breeds, together with strong selection for traits, has led to the establishment of a unique resource for dissecting the genetic basis of simple and complex mammalian traits. Toward this end, we present a comprehensive radiation hybrid map of the canine genome composed of 3,270 markers including 1,596 microsatellite-based markers, 900 cloned gene sequences and ESTs, 668 canine-specific bacterial artificial chromosome (BAC) ends, and 106 sequence-tagged sites. The map was constructed by using the RHDF5000-2 whole-genome radiation hybrid panel and computed by using MULTIMAP and TSP/CONCORDE. The 3,270 markers map to 3,021 unique positions and define an average intermarker distance corresponding to 1 Mb. We also define a minimal screening set of 325 highly informative well spaced markers, to be used in the initiation of genome-wide scans. The well defined synteny between the dog and human genomes, established in part as a function of this work by the identification of 85 conserved fragments, will allow follow-up of initial findings of linkage by selection of candidate genes from the human genome sequence. This work continues to define the canine system as the method of choice in the pursuit of the genes causing mammalian variation and disease.

dog | microsatellites | ESTs | bacterial artificial chromosome ends

The structure of the canine population offers unparalleled opportunities for understanding the genetic basis of morphology, behavior, and disease susceptibility (1–3). Millions of purebred dogs are newly registered worldwide every year, each of which will be assigned to one of ≈300 well defined "breeds" based on its parentage (4). To maintain physical and behavioral homogeneity, gene flow between breeds is tightly restricted and only a dog whose parents are both registered members of a breed is also eligible for registration.

Global events including world wars and economic depressions have limited the number of founders, and thus restricted the genetic diversity associated with many dog breeds. This, together with the common practice of repeatedly breeding dogs that feature desired physical or behavioral characteristics has resulted in severe population bottlenecks within many breeds, at times reducing the effective breeding stock to only a few individuals (5). The net result is a species characterized by enormous phenotypic diversity, but often at a loss of genome-wide variability (5). As a result, inherited diseases are common in most dog breeds. Researchers concerned with human disease gene mapping are thus afforded a rare opportunity to understand the genetics of diseases that have proven intractable through the study of small, outbred human families (3, 6, 7). In addition, the phenotypic diversity present in modern dog breeds offers developmental biologists an opportunity to decipher the contributions of multiple interacting loci to the seemingly complex phenotypes associated with mammalian development (8). Toward that end, we have developed the resources for mapping and sequencing the dog genome (9–11). Our most recent efforts, summarized herein, encompass a complete mapping resource featuring a 3,270-marker radiation hybrid (RH) map that spans the entire dog genome at 1-Mb resolution, with a well distributed set of microsatellite markers, mapped bacterial artificial chromosome (BAC) ends, and canine-specific genes or ESTs.

Methods

Genotyping. The panel used in these experiments, RHDF5000-2, comprises 118 cell lines from the original RHDF5000 panel, constructed by fusing dog fibroblasts irradiated at 5,000 rad with TK-HTK3 hamster cells (12). The panel has a retention frequency of 22% with a theoretical resolution limit of 600 kb.

Primers were designed to have an optimal length of 25 nt and a melting temperature of 58–60°C, and result in amplicons of 200–500 bp. PCRs were carried out in 15-μl volumes as described (9–11) by using the following touchdown program: 8 min 95°C, followed by 20 cycles of 30 sec 94°C, 30 sec 63°C decreasing of 0.5°C per cycle, 1 min 72°C and 15 cycles of 30 sec 94°C, 30 sec 53°C, 1 min 72°C, and a final extension of 2 min 72°C. Markers yielding either faint or spurious bands were optimized. Amplification products were resolved and recorded as described (9, 11). Accession numbers, characterization, and PCR conditions for all markers are available at www-recomgen.univ-rennes1.fr/doggy.html and www.fhcrc.org/science/dog_genome/dog.html.

Microsatellite Markers. New microsatellite markers were isolated and characterized as described (13). The degree of polymorphism was estimated either as a heterozygosity (Het) value or a polymorphic information content (PIC) value after testing a panel of either 5 unrelated mongrel dogs (14) or 10 unrelated purebred dogs representing a subset of the 20 most popular American Kennel Club breeds (13).

BAC-End Sequences. Plates of BAC clones were randomly selected from the RPC81 canine BAC library (15) for end-sequencing using standard automated approaches (16). Average read lengths were in excess of 700 bp. Primers defining each BAC end were selected from sequence with the highest number of high-quality (HQ) bases. HQ sequence was defined as having 100 continuous sequences with PHRED scores of 20 or greater. Only one set of primers was used to genotype each BAC; primers designed from the opposite end of the insert were used for genotyping only if the first pair yielded poor-quality data.

Single-Nucleotide Polymorphism (SNP)-Containing Sequence-Tagged Site (STS) Markers. A genomic library was constructed by cloning 1-kb inserts of mongrel dog genomic DNA into pBluescript KS+II

Abbreviations: BAC, bacterial artificial chromosome; CS, conserved segments; RH, radiation hybrid; TSP, traveling salesman problem; SNP, single-nucleotide polymorphism; FISH, fluorescence *in situ* hybridization; STS, sequence-tagged sites; CFA, *Canis familiaris*; HSA, *Homo sapiens*; Het, heterozygosity; lod, logarithm of odds.

‡Present address: Stanford University School of Medicine, Center For Narcolepsy, 701B Welch Road, Palo Alto, CA 94305.

§Present address: Regulome Corporation, 551 North 34th Street, Seattle, WA 98103.

¶To whom correspondence may be addressed. E-mail: galibert@univ-rennes1.fr or eostrand@fhcrc.org.

vector. Two hundred clones were sequenced and SNPs were identified after analysis of STS using DNA isolated from 20 dogs representing 20 breeds. Cycle sequencing was performed by using the BigDye Terminator chemistry (Applied Biosystems). Sequencing traces were processed by using PHRED, PHRAP, and CONSED (17–19). SNPs were identified by visual inspection of mismatches detected in the 20 sequencing traces.

Gene and EST-Based Markers. Primers were designed to amplify known dog gene sequences deposited in public databases by using standard approaches (9, 11). Canine ESTs were isolated from a cDNA library constructed from a Madin–Darby canine kidney cells line by priming with a tailed oligo (dT).

Identification of Orthologous Human Gene Sequences. Orthologous human sequences were searched by BLAST analyses (20) against public databases (GenBank "nr" and "HTGS") by using default criteria and by BLAT searches (21) against "Human NCBI build 31" sequence. For 95% of the genes, a sequence analogy >80% over 100 nt was observed. The size of 100 bp for sequence comparison was dictated by the size of the available query sequence and not by the absence of analogy. Gene nomenclatures were retrieved from the LocusLink database and human chromosomal locations were confirmed by the University of California Santa Cruz human genome server (Nov. 2002); http://genome.ucsc.edu.

Quality Control. Approximately 65% of BAC end markers and 30% of gene-based and microsatellite markers were genotyped in duplicate. These correspond to a subset of markers selected at random, as well as to gene markers mapping to regions of synteny breaks. Additional markers were selected from RH groups where ambiguities in ordering were noted and all singletons were also typed in duplicate. Duplicate data were considered consistent when the number of discrepancies between data sets was ≤16%. The percent was calculated as the number of differences over the marker retention value. A threshold limit was determined as corresponding to a distance lower than the resolution limit of the RHDF5000-2 panel. In rare cases, where two independent typings yielded >16% discrepancies, a third typing was done and the resulting vector was either integrated into the map construction, or the marker was discarded if no agreement was observed between two of three genotypes.

Analysis and Map Construction. Novel markers were incorporated into the previous 1,500-marker RH data set (11) by pairwise calculations using MULTIMAP software (22) at a logarithm of odds (lod) threshold ≥8.0. A total of 3,162 markers could be clustered into RH groups. RH groups were ordered by using the traveling salesman problem (TSP) approach as specified by the CONCORDE computer package (23). TSP/CONCORDE computes five independent RH maps; three are variants of the maximum-likelihood estimate approach, and two are constructed by using obligate chromosome breaks. The resulting maps were evaluated to produce a consensus map (24). For markers whose map position was not well supported, genotyping data were reexamined, and genotypes were repeated. When no erroneous genotypes were observed, the problematic linkage group was split into two or more RH groups by using the MULTIMAP algorithm and a lod threshold of >9.0.

Inter-marker distances were determined with the rh_tsp_map1.0 version of TSP/CONCORDE, which delivers map positions in arbitrary units. For each chromosome the sum of the arbitrary units was converted into kb by using the known physical size of each chromosome, as determined by cytofluorimetry (25). When more than one RH group was assigned to a chromosome, 350 units were added for each gap, corresponding to the upper limit of our ability to detect linkage between adjacent markers.

Results

General Map Characteristics. The 1,770 markers added to the canine RH map were typed on the RHDF5000-2 panel described (11, 26). Mapping vectors were added to the previous 1,500-marker map data set (11), and the complete dataset of 3,270 markers was recomputed by using MULTIMAP (22) and TSP/CONCORDE (23) software programs. Pairwise linkage analysis at a lod threshold ≥8.0 by using MULTIMAP allowed the localization of 3,162 markers to the 38 autosomes and sex chromosomes, leaving only 16 orphan RH groups and 108 unlinked markers. Of the 16 orphan groups, comprising 2–19 markers, 12 could be incorporated into RH groups already assigned to chromosomes by using two-point analyses with lod scores between 5.0 and 8.0. For eight groups the resulting map position is in full agreement with predictions from syntenic human data, and for one group a synteny break is introduced. The four remaining orphan RH groups contain only 14 markers.

Ordering of markers within each RH group was performed by using the TSP/CONCORDE software (23). The number of markers assigned to each autosome ranged from 156 markers at 147 unique positions on chromosome 1 (*Canis familiaris*, CFA 1) to a minimum of 25 markers at 24 positions (CFA 38). The smallest canine chromosome, the Y, has 10 markers (Table 1). TSP/CONCORDE (23) provides distances between markers in arbitrary units. For each chromosome, we converted the sum of the arbitrary units into kb, with a mean value of 1 unit corresponding to 11.8 kb, as calculated from a subset of well covered chromosomes (Table 1).

The total map size for individual autosomes ranges from 12,353 units (CFA 1) to 1,783 units (CFA 38) (Table 1). The total size of the complete RH map is 227,127 units. The 3,270 markers map to 3,021 unique positions; 249 markers (8%) are copositioned. In one case, CFA 35, five independent markers colocalize to a unique position. The average intermarker distance of the map is 78 units, or ≈900 kb. The present map, therefore, represents a global 2-fold increase in marker density compared with previous iterations of the map (11), with a concomitant 1.5-fold increase in the number of microsatellite markers, a 2.8-fold increase in EST/gene markers and a novel set of mapped BAC end sequences. With this current data set of markers the RHDF5000-2 panel has yet to reach saturation; the resolution of the resulting canine RH map, however, now stands at <1 Mb.

Map Coverage. We used a variety of different methods to estimate a coverage of 90–95% for the previously reported 1,500-marker RH map (11). In the present effort we have more than doubled the number of markers on the map and, as expected, significantly better genome coverage is now attained. By taking advantage of the fact that some markers placed on the RH map were previously localized by fluorescence *in situ* hybridization (FISH) (11), we conclude that coverage is now complete or nearly complete for most chromosomes. This is easiest to ascertain when markers corresponding to FISH probes localized to telomeres were then found to map to the extremities of RH groups, or when additional markers were mapped between a FISH probe and a telomeric end; i.e., CFA 34, where six markers were added to the terminal portion of the RH group, and CFA 10 and 23. For chromosomes with complete coverage, one arbitrary unit corresponds to 10–15 kb. We do note, however, that coverage is not absolutely complete for some chromosomes. For instance, comparison of RH and FISH mapping data suggests that CFA 32 and 35 are covered by smaller RH groups than expected; for those chromosomes the arbitrary unit corresponds to 21 and 16 kb, respectively. Also, in the case of CFA 5, we know that a region including and proximal to the *p53* gene was not retained when the hybrid lines were constructed (9, 27). In the case of CFA 13 and 38, the number of marker positions (59 and 24, respectively)

Table 1. Map statistics by chromosome

CFA	CFA size*, Mb	RH map size, TSP units	Ratio, TSP units/kb	Total no. of positions	Intermarker average distance†, Mb	Total no. of markers	No. of microsatellite markers	No. of gene-based markers	No. of BAC-end markers	No. of STS markers‡	No. of CFA-specific markers§	No. of Zoo-FISH CS¶	No. of RH CS‖	No. of singletons**
1	137	12,353	11	147	0.93	156	79	29	42	6	2	4	4	3
2	99	8,233	12	105	0.94	114	60	29	20	5	3	4	3	1
3	105	9,165	11	112	0.94	116	60	23	27	6	3	3	3	
4	100	7,587	13	100	1.00	115	52	33	23	7	3	3	3	
5	99	7,896	13	111	0.89	123	60	30	30	3	7	4	4	1
6	87	6,056	14	84	1.04	99	41	37	19	2	6	3	3	
7	94	6,834	14	116	0.81	136	53	46	31	6	3	2	2	
8	86	6,399	13	87	0.99	92	39	34	18	1	2	1	1	
9	77	7,557	10	100	0.77	111	40	48	16	7	6	2	2	
10	80	6,722	12	87	0.92	96	43	27	21	5	3	3	3	
11	86	7,257	12	104	0.83	107	53	23	25	6	3	2	2	
12	85	9,002	9	119	0.71	133	73	34	23	3	2	1	1	
13	75	4,392	17	59	1.27	68	36	20	11	1	1	2	2	
14	72	6,867	10	80	0.90	86	39	22	22	3	2	1	2	
15	75	7,523	10	81	0.93	82	43	22	13	4	5	3	5	
16	73	6,863	11	65	1.12	69	34	18	17	0	1	2	3	
17	80	6,081	13	85	0.94	93	47	32	12	2	3	2	2	
18	66	6,967	9	82	0.80	89	40	29	17	3	4	2	2	
19	66	4,468	15	60	1.10	70	50	8	11	1	2	2	2	
20	66	4,219	16	98	0.67	107	49	35	19	4	5	2	2	
21	61	8,045	8	93	0.66	95	49	29	16	1	1	1	1	1
22	61	5,349	11	74	0.82	82	51	15	16	0	2	1	1	
23	61	5,382	11	61	1.00	62	31	13	17	1	3	1	1	
24	73	5,341	14	57	1.28	63	31	12	17	3	2	1	1	
25	60	5,833	10	70	0.86	72	32	19	17	4	2	3	4	
26	48	3,256	15	53	0.91	57	28	18	11	0	2	2	2	1
27	57	7,116	8	72	0.79	78	39	30	8	1	2	1	1	1
28	55	3,332	17	58	0.95	61	27	18	14	2	2	1	1	
29	51	5,230	10	59	0.86	59	39	6	9	5	2	1	1	
30	47	3,740	13	45	1.04	48	21	14	10	3	2	1	1	
31	50	2,779	18	37	1.35	38	19	9	8	2	2	2	2	
32	51	2,477	21	29	1.76	30	15	6	8	1	1	1	1	
33	41	3,561	12	45	0.91	49	19	16	14	0	1	1	1	
34	50	3,863	13	51	0.98	58	33	10	13	2	2	1	2	
35	38	2,406	16	28	1.36	33	20	6	7	0	1	1	1	
36	41	3,968	10	46	0.89	47	19	16	11	1	1	1	1	
37	40	2,877	14	48	0.83	52	29	10	12	1	2	1	1	
38	38	1,783	21	24	1.58	25	9	8	8	0	1	1	1	
X	139	5,225	25	53	2.62	59	25	20	14	0	1	1	1	
Y	27	1,714	10	10	2.70	10	8	1	1	0	1	0	0	1
Total assigned		225,718		2,895		3,140	1,535	855	648	102	99	71	76	9
Average			11,8††		0.96									
Orphan groups		1,409		20		22	15	0	6	1	0			
Unlinked				106		108	46	45	14	3	0			
Total	2,797	227,127		3,021		3,270	1,596	900	668	106	99	71	76	9

*Chromosome sizes are given in Mb from cytofluorimetry measurements (25).
†Average intermarker distances (in Mb) are calculated by dividing the size of the chromosome by the number of unique positions.
‡SNP-containing STS and CFA-specific STS markers.
§Markers derived from clones for FISH experiments in Breen *et al.* (11). These markers are included in other marker categories and are not counted in the total number of markers.
¶Zoo-FISH CS refer to human/dog conserved segments identified by Zoo-FISH data (30, 31).
‖Human/dog conserved segments identified from the RH map; a CS comprises two or more markers.
**Putative CS identified by RH mapping but containing only one marker.
††This value is calculated from the subset of well covered chromosomes (all but CFA5, 32, 35, 38, X, and Y).

appears low considering the size of each chromosome (75 and 38 Mb, respectively). Consequently, either marker density is low for these chromosomes and/or coverage is incomplete. The presence of only a single FISH marker located near the middle of the chromosome does not allow us to distinguish between these possibilities.

Despite its large size, a relatively small number of markers have been placed on the canine X chromosome, which can be partly explained by the reported paucity of genes on mammalian X chromosomes. Thus, the existence of several unlinked RH groups of unknown spacing on the X chromosome is not surprising and reported distances probably underestimate true

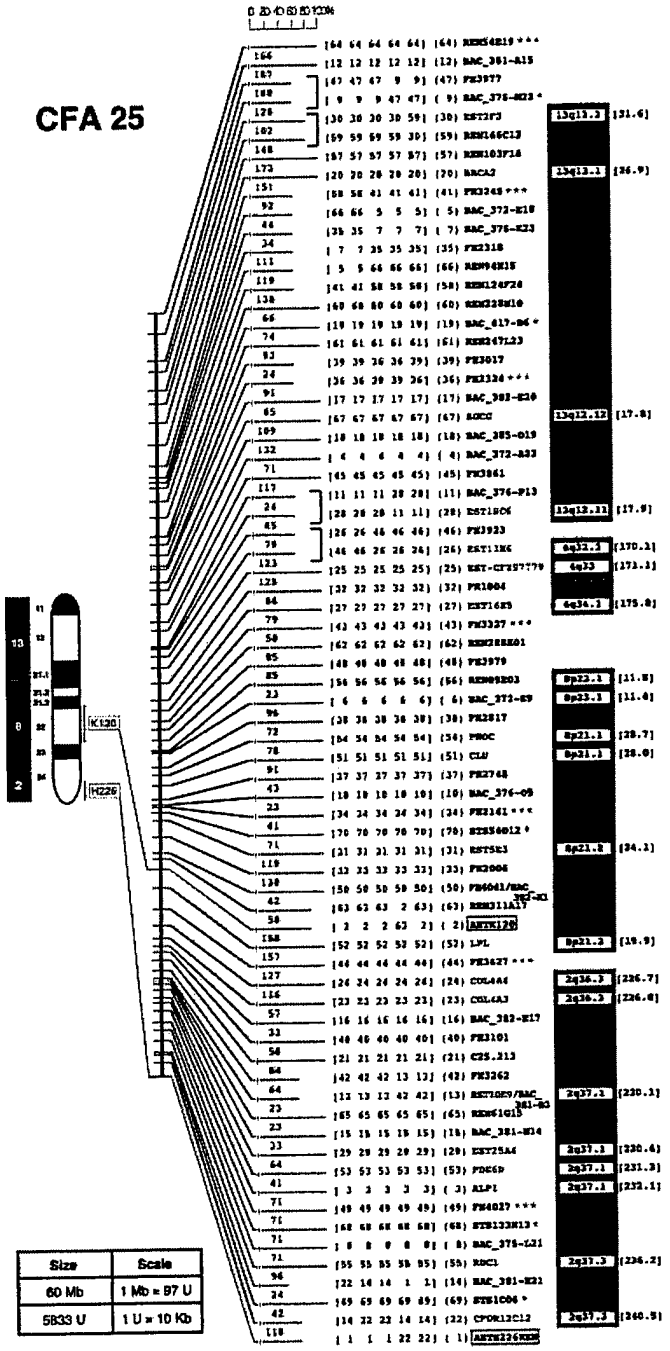


Fig. 1. RH map of CFA25. The position of each marker is reported along the RH map, symbolized by a vertical bar. The RH map shows the five maps generated by TSP/CONCORDE. Maps are highlighted by horizontal bars of variable lengths. When a marker is present on all five maps at the same position, the horizontal bar has a maximum length indicating high confidence; shorter bars reflect a lower confidence level. The scale of 0–100% reflects the relative confidence level. Marker number, as it appears in the consensus map, is indicated in parentheses. In scrambled regions, markers occupying several positions are bracketed to narrow the problematic region into smaller intervals. Marker names indicated in red correspond to gene-based markers (type I); other markers are black (see Table 1 for nomenclature). MSS-2 markers have three asterisks; polymorphic STS, genes, or BAC ends have one. Colored boxes to the right of the markers display human segments with the chromosomal band position. The corresponding position in nucleotide (Mb) of human putative orthologs is indicated between brackets. Data are based on NCBI Build 31. At the left of the RH map, a 4′,6-diamidino-2-phenylindole-banded ideogram is drawn. Markers assigned to chromosomes by FISH are linked to their RH map positions by colored lines (11). Colored bars correspond to the human evolutionary CS. Numbers indicate HSA

interval size. We did investigate use of a lod threshold of 6.0 rather than 8.0 to see whether the overall X map could be improved. That adjustment does result in generation of two large linkage groups, rather than seven obtained when a lod of 8.0 was used. But the ordering of these two groups was suboptimal and only the map constructed at lod 8.0 is presented.

Microsatellite Characteristics. In addition to the previously placed 1,078 microsatellites, 518 microsatellite based markers have been added to the map and a total of 1,005, 20, and 571 microsatellite markers based on di-, tri-, and tetranucleotide repeats, respectively, are now mapped. Markers are randomly distributed throughout the chromosomes, ranging from the fewest (9 on CFA 38) to the most (79 on CFA 1). Polymorphism was evaluated by estimation of Het and/or PIC values for markers with 12 or more repeat units. Of these, 77% had Het or PIC values >0.5 and 480 had values >0.7. Because polymorphism levels have not been assessed for every marker on the map, the actual number is likely to be higher.

Minimal Screening Set of Microsatellite-Based Markers for Genome-Wide Scans. We developed a minimal screening set (MSS-2) of 325 markers with an average spacing of 9 Mb to be used for genome-wide scans in the dog. Criteria for marker selection, in order of preference, included spacing (interval distribution >800 kb and <12,000 kb), informativeness, cleanliness of PCR product, and amplicon size. Preference was given to markers generating PCR products <500 bp. When possible, for chromosomes in which multiple RH groups were present, markers were selected that defined the ends of each RH group. Markers mapping to CFA Y were also selected, as they may prove useful for forensic studies, paternity testing, and for defining pseudoautosomal regions on the sex chromosomes. The final minimal screening set spans 81 RH groups and all chromosomes. The average Het is 0.73, with 171 tetra-, 151 di-, and 3 markers based on trinucleotide repeats. The largest known interval, located on CFA 8 between FH3241 and REN204K13, is 17.1 Mb. Fifty-six markers were also part of the MSS-1 set (28).

A Framework of RH Mapped BAC Clones. From a selected set of 2,016 BACs we obtained high-quality sequences from either one or both ends of 1,504 BACs (766 for one end only, 738 for both ends). The 4,032 sequences generated had an average of 342 bases with PHRED scores ≥20. Markers were designed for several hundred clones, and 668 have now been genotyped across the RHDF5000-2 panel. BAC ends are randomly distributed throughout all chromosomes; ranging from one on CFA Y to 42 on CFA 1. These 668 mapped BAC ends constitute an initial framework of clones for anchoring the canine physical map and provide a format for positional cloning studies. A subset of 39 mapped BAC clones also contained microsatellites within the end sequences. These are indicated in Fig. 1 and all associated figures found at www-recomgen.univ-rennes1.fr/doggy.html and www.fhcrc.org/science/dog_genome/dog.html.

STS-Containing SNP Markers. A total of 200 STS were isolated and sequenced from a canine genomic DNA library. Seventy-eight SNPs were found by sequencing each STS in 20 dogs belonging to different breeds and 72 STSs, containing one to six SNPs, were

origin as determined by reciprocal chromosome painting (30, 31). Distances between RH markers are reported in TSP units between horizontal bars. The physical size of each chromosome (in Mb), as determined by flow sorting (25), and the RH group total size (in TSP units) are reported in the frame. The correspondences between TSP unit and kb are also reported in the frame. Figures for all chromosomes are available at www-recomgen.univ-rennes1.fr/doggy.html and www.fhcrc.org/science/dog_genome/dog.html.

RH mapped. These are distributed on 29 of 38 canine autosomes. Relevant characteristics including sequence context, minor allele frequency, and heterozygosity can be found at www-recomgen.univ-rennes1.fr/doggy.html and www.fhcrc.org/science/dog_genome/dog.html. These polymorphic markers are indicated by a star in Fig. 1 and all figures found at the web sites listed above.

Gene-Based Markers and Comparative Mapping. A total of 900 gene based markers were incorporated into the present version of the map, of which 580 are novel, representing a 2.8-fold increase over that presented previously (11). Four hundred forty-one represent novel ESTs for which localization of the human ortholog is known. The remaining 139 are canine gene markers of diverse origins (see the table on the web sites cited above). Some have been shown previously to be polymorphic (29) as indicated by a star in Fig. 1. The distribution of gene-based markers averages one per 3 Mb, with such markers now well distributed across all chromosomes (Table 1). CFA 32 and CFA 36, which lacked any gene-based markers on the previous map (11) now contain 6 and 16 mapped ESTs, respectively.

From the total set of 900 markers, 820 have a known orthologous localization in the human genome. This provides 780 unique positions for comparison with the human genome map. For BAC ends, microsatellites, and STS markers located in regions between conserved segments, the sequences of the original clones were tested by BLAT searches against the Human "NCBI Build 31" sequence. Of 380 sequences, 50 (13%) gave reliable localizations. Thus, a total of 870 canine mapped sequences occupying 830 unique positions have a known human localization, allowing anchorage of the canine and human genomes.

The mapping of these 870 markers allows us to confirm all but one of the conserved segments (CS) detected by human-on-dog chromosome paint studies (30, 31), or those previously identified by RH mapping as singletons (fragment containing only one gene) (11). Only the human chromosome 19 (*Homo sapiens*, HSA19p13) singleton containing UBA52 was discarded during the present RH map construction. Moreover, five novel CS, containing between two and four mapped genes, all with a high level of sequence analogy with their human counterparts (see *Methods*) have been detected: CFA14/HSA1, CFA15/HSA14 and HSA16, CFA25/HSA4, and CFA34/HSA5. In addition, five novel singletons (CFA1/HSA8, HSA4, HSA22; CFA5/HSA2; CFA21/HSA15) sharing a high level of sequence identity with their human counterpart (>91% for more than 190 nt) and two with a lower support CFA26/HSA10 (86% over 1,148 nt) and CFA27/HSA18 (85% over 139 nt) are detected. Until other mapped genes confirm their status as CS, the singletons should be interpreted with caution. We believe they are likely to be correct, however, as 16 of the 18 singletons detected previously by using the same criteria (11) have been confirmed by RH mapping of additional markers as conserved segments in this study. In total, therefore, 85 human/dog orthologous fragments corresponding to 76 CS plus 9 singletons, are presently observed by RH mapping (Fig. 2).

Conserved syntenic fragments between dog and human are shown for CFA25 on Fig. 1 and illustrated in Fig. 2. A total of 16 dog chromosomes appear to correspond to only one human fragment (CFA8 = most of HSA14q; CFA12 = most of HSA6; CFA22–24, 28–30, 32, 33, and 35–38 plus X and Y). The 24 remaining correspond to between two and seven unique human chromosomal fragments (singletons included) (Fig. 2). Only one human autosome, HSA20, shares exclusive synteny with a unique dog chromosome, CFA24. Gene order at G-banding resolution is also conserved. All other human chromosomes contain from two to nine conserved canine segments with HSA1 containing most. In addition, the size of most previously described chromosomal segments are now substantially extended. Consider, for instance, CFA3 (limits between HSA15 and HSA4) and CFA6 (limits between HSA16 and HSA1) or CFA25, where the limits between human conserved segments HSA13, HSA4, HSA8, and HSA2 are more accurately defined (Fig. 1).

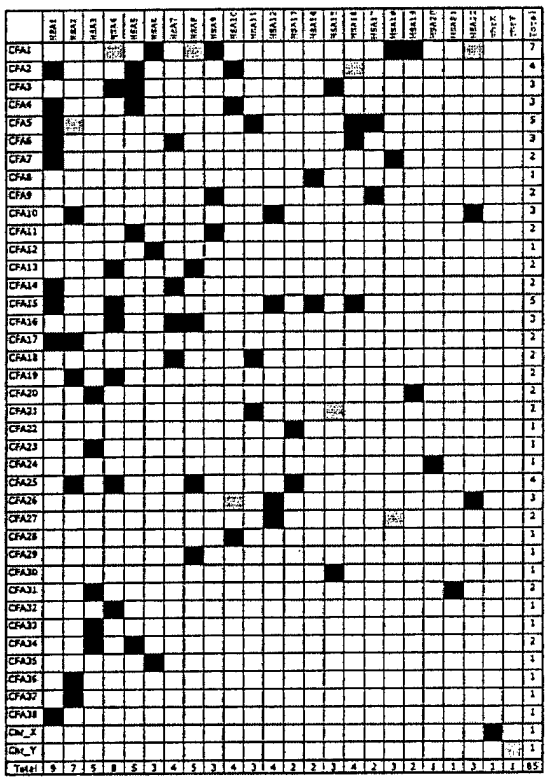


Fig. 2. Schematic view of RH conserved segments and singletons between dog and human. CS between both species are illustrated by black squares; singletons are illustrated by gray squares. For each CFA and HSA, the total number of CS is reported in the last column and the last line, respectively.

Discussion

Significant progress has been made in the development of the canine genetic system recently (9–11, 32). In recent years we, and others, have demonstrated the genetic power of canines by mapping and/or cloning several disease genes, as summarized in our White Paper Proposal for Sequencing the Canine Genome (www.genome.gov/page.cfm?pageID=10002154). This has led to an increased utilization of the canine system for the development of gene therapy protocols (33–35) or *in vivo* targeted repair (36). Moreover, the utilization of the map to identify quantitative trait loci appears promising, as demonstrated by the recent study identifying loci for canine morphology and development (8).

This most recent iteration of the map features three major advances: (*i*) the presentation of a second minimal screening set of markers to be used for undertaking genome-wide scans; (*ii*) the placement of an initial set of BAC end sequences to facilitate positional cloning studies; and (*iii*) refinement of the canine/human comparative map.

The first advance featured herein is the presentation of a well characterized minimal screening set of markers (MSS-2) for undertaking genome-wide scans. The density and overall informativeness of this set surpasses that presented previously; the overall Het values are higher, and the coverage across the genome is, at a minimum, 25% denser (28). If we consider that the 325 markers define 253 intervals of known size within RH groups, only 21 of those are ≥12 Mb and a majority (166) are ≥8

Mb and ≤12 Mb. The smallest intervals appear at the ends of radiation hybrid groups, where additional markers were picked to ensure that areas bordering unknown distances beyond RH groups were appropriately covered.

One issue of ongoing concern is the degree to which any set of starting markers will be useful for genome scans in purebred dogs. Some breeds appear as outbred as the general human population, whereas others, because of popular sire effects, bottlenecks, and selective breeding, display limited genetic heterogeneity (5). A key task for the future is the development of markers that are polymorphic in multiple breeds.

A second major advance in the current map is the initial placement of a large set of BAC end sequences. Although this initial data set includes only 668 mapped BAC ends, the resultant density is sufficient that any mapped locus is likely to be close enough to multiple BACs that the construction of minimum tiling paths can be initiated.

The final major advance is summarized by the now detailed information available regarding evolutionary relationships between the human and canine genomes. The First International Workshop on Comparative Genome Organization has suggested several levels of conservation to consider when comparing genomes of two different species. The most relevant at this point in map development are conserved segments, i.e., the syntenic association of two or more contiguous, homologous genes in separate species (37). Previous human-on-dog chromosome painting studies identified 68 conserved chromosome segments, including the X chromosome (30), or 73 excluding the X (31). Conversely, 90 independent segments were identified with dog-on-human chromosome paints (31, 38). The present work is comparable in principle to human-on-dog chromosome paints and the number of conserved segments presented here are best compared with the 68 and 73 reported by Breen *et al.* (30) and Yang *et al.* (31). The analysis presented here allowed us to identify all but one previously detected conserved segments (30, 31). In addition we detect 12 novel orthologous fragments, i.e., five chromosomal segments and seven singletons (Table 1). In total, therefore, 85 human/dog orthologous fragments, 76 CS plus 9 singletons, are presently observed by RH mapping (Fig. 2).

When considering the conservation of gene order between human and dog at the human G banding level, for CS harboring more than 10 genes, three interesting situations are observed: (*i*) CS in which the gene order is very well conserved between the two species, i.e., CFA8/HSA14; CFA12/HSA6, CFA27/HSA12, CFA30/HSA15, CFA33/HSA3, and CFA36/HSA2. (*ii*) CS that can be split into several blocks where gene order is conserved. This is the case for CFA 4/HSA5, CFA14/HSA7 CFA17/HSA2, CFA21/HSA11, CFA22/HSA13, and CFA24/HSA20. This is often observed when the human orthologous segments span the centromeres. (*iii*) CS in which the gene order is not conserved, as for CFA 9/HSA17. To more precisely map such CS, denser gene maps built with higher-resolution RH panels will be needed.

Finally, this work highlights the utility and major advantages of using the TSP/CONCORDE algorithm. Recalling that RH maps are not physical maps *per se*, but rather based on a statistical treatment of a set of mapping vectors, the TSP/CONCORDE algorithm allows an unbiased representation of the data, rather than favoring any single interpretation. In addition, by assigning a level of confidence with which each marker can be assigned to a given position, map users can more appropriately adapt cloning strategies to fit specific needs. Recombination intervals defined by markers positioned with high confidence can reduce the overall workload associated with building a physical map across a region of interest. BACs and ESTs mapped with a high degree of confidence facilitate orientation of the map with the corresponding region of the human genome. The work presented here, therefore, provides a refined set of resources for using comparative approaches to map and clone genes of interest in the canine system.

We acknowledge the American Kennel Club Canine Health Foundation, the Burroughs Wellcome Fund, U.S. Army Grant DAAD19-01-1-0658, and National Institutes of Health Grant R01CA-92167 (to E.A.O. and F.G.). R.G. is partly supported by an American Kennel Club and Centre National de la Recherche Scientifique fellowship, and P.Q. is partly supported by a Conseil Regional de Bretagne fellowship. L.K. is supported by a Nestle Purina fellowship, H.G.P. is supported by Public Health Service Grant T32 HG00035, and J.K.L. is supported by Public Health Service National Research Service Award T32 GM07270 from the National Institute of General Medical Sciences.

1. Ostrander, E. A. & Giniger, E. (1997) *Am. J. Hum. Genet.* **61**, 475–480.
2. Ostrander, E. A., Galibert, F. & Patterson, D. F. (2000) *Trends Genet.* **16**, 117–123.
3. Patterson, D. (2000) *J. Vet. Intern. Med.* **14**, 1–9.
4. Wilcox, B. & Walkowicz, C. (1995) *Atlas of Dog Breeds of the World* (T.H.F., Neptune City, NJ).
5. Ostrander, E. A. & Kruglyak, L. (2000) *Genome Res.* **10**, 1271–1274.
6. Patterson, D. F. (2002) *Canine Genetic Disease Information System: A Computerized Knowledgebase of Genetic Diseases in Dogs* (Mosby-Elsevier, St. Louis).
7. Galibert, F., Andre, C., Cheron, A., Chuat, J. C., Hitte, C., Jiang, Z., Jouquand, S., Priat, C., Renier, C. & Vignaux, F. (1998) *Bull. Acad. Natl. Med.* **182**, 811–821.
8. Chase, K., Carrier, D. R., Adler, F. R., Jarvik, T., Ostrander, E. A., Lorentzen, T. D. & Lark, K. G. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 9930–9935.
9. Priat, C., Hitte, C., Vignaux, F., Renier, C., Jiang, Z., Jouquand, S., Cheron, A., Andre, C. & Galibert, F. (1998) *Genomics* **54**, 361–378.
10. Mellersh, C. S., Hitte, C., Richman, M., Vignaux, F., Priat, C., Jouquand, S., Werner, P., André, C., DeRose, S., Patterson, D. F., *et al.* (2000) *Mamm. Genome* **11**, 120–130.
11. Breen, M., Jouquand, S., Renier, C., Mellersh, C. S., Hitte, C., Holmes, N. G., Cheron, A., Suter, N., Vignaux, F., Bristow, A. E., *et al.* (2001) *Genome Res.* **11**, 1784–1795.
12. Vignaux, F., Hitte, C., Priat, C., Chuat, J. C., Andre, C. & Galibert, F. (1999) *Mamm. Genome* **10**, 888–894.
13. Ostrander, E. A., Sprague, G. F., Jr., & Rine, J. (1993) *Genomics* **16**, 207–213.
14. Jouquand, S., Chéron, A. & Galibert, F. (1999) *Biotechniques* **26**, 902–905.
15. Li, R., Mignot, E., Faraco, J., Kadotani, H., Cantanese, J., Zhao, B., Lin, X., Hilton, L., Ostrander, E. A., Patterson, D. F., *et al.* (1999) *Genomics* **58**, 9–17.
16. Mahairas, G. G., Wallace, J. C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., *et al.* (1999) *Proc. Natl. Acad. Sci. USA* **96**, 9739–9744.
17. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. (1998) *Genome Res.* **8**, 175–185.
18. Ewing, B. & Green, P. (1998) *Genome Res.* **8**, 186–194.
19. Gordon, D., Abajian, C. & Green, P. (1998) *Genome Res.* **8**, 195–202.
20. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402.
21. Kent, W. J. (2002) *Genome Res.* **12**, 656–664.
22. Matise, T. C., Perlin, M. & Chakravarti, A. (1994) *Nat. Genet.* **6**, 384–390.
23. Agarwala, R., Applegate, D. L., Maglott, D., Schuler, G. D. & Schaffer, A. A. (2000) *Genome Res.* **10**, 350–364.
24. Hitte, C., Lorentzen, T. D., Guyon, R., Kim, L., Cadieu, E., Parker, H., Quignon, P., Lowe, J., Gelfenbeyn, B., Andre, C., *et al.* (2003) *J. Hered.*, in press.
25. Langford, C. F., Fischer, P. E., Binns, M. M., Holmes, N. G. & Carter, N. P. (1996) *Chromosome Res.* **4**, 115–123.
26. Vignaux, F., Priat, C., Jouquand, S., Hitte, C., Jiang, Z., Cheron, A., Renier, C., Andre, C. & Galibert, F. (1999) *J. Hered.* **90**, 62–67.
27. Jonasdottir, T. J., Mellersh, C. S., Moe, L., Heggebo, R., Gamlem, H., Ostrander, E. A. & Lingaas, F. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 4132–4137.
28. Richman, M., Mellersh, C. S., Andre, C., Galibert, F. & Ostrander, E. A. (2001) *J. Biochem. Biophys. Methods* **47**, 137–149.
29. Parker, H. G., Yuhua, X., Mellersh, C. S., Khan, S., Shibuya, H., Johnson, G. S. & Ostrander, E. A. (2001) *Mamm. Genome* **12**, 713–718.
30. Breen, M., Thomas, R., Binns, M. M., Carter, N. P. & Langford, C. F. (1999) *Genomics* **61**, 145–155.
31. Yang, F., O'Brien, P. C., Milne, B. S., Graphodatsky, A. S., Solanky, N., Trifonov, V., Rens, W., Sargan, D. & Ferguson-Smith, M. A. (1999) *Genomics* **62**, 189–202.
32. Mellersh, C. S., Langston, A. A., Acland, G. M., Fleming, M. A., Ray, K., Wiegand, N. A., Francisco, L. V., Gibbs, M., Aguirre, G. D. & Ostrander, E. A. (1997) *Genomics* **46**, 326–336.
33. Acland, G. M., Aguirre, G. D., Ray, J., Zhang, Q., Aleman, T. S., Cideciyan, A. V., Pearce-Kelling, S. E., Anand, V., Zeng, Y., Maguire, A. M., *et al.* (2001) *Nat. Genet.* **28**, 92–95.
34. Ponder, K. P., Melniczek, J. R., Xu, L., Weil, M. A., O'Malley, T. M., O'Donnell, P. A., Knox, V. W., Aguirre, G. D., Mazrier, H., Ellinwood, N. M., *et al.* (2002) *Proc. Natl. Acad. Sci. USA* **99**, 13102–13107.
35. Herzog, R. W., Yang, E. Y., Couto, L. B., Hagstrom, J. N., Elwell, D., Fields, P. A., Burton, M., Bellinger, D. A., Read, M. S., Brinkhous, K. M., *et al.* (1999) *Nat. Med.* **5**, 56–63.
36. Bartlett, R. J., Stockinger, S., Denis, M. M., Bartlett, W. T., Inverardi, L., Le, T. T., thi Man, N., Morris, G. E., Bogan, D. J., Metcalf-Bogan, J. & Kornegay, J. N. (2000) *Nat. Biotechnol.* **18**, 615–622.
37. Andersson, L., Archibald, A., Ashburner, M., Audun, S., Barendse, W., Bitgood, J., Bottema, C., Broad, T., Brown, S., Burt, D., *et al.* (1996) *Mamm. Genome* **7**, 717–734.
38. Sargan, D. R., Yang, F., Squire, M., Milne, B. S., O'Brien, P. C. & Ferguson-Smith, M. A. (2000) *Genomics* **69**, 182–195.

J Hered. 94:9-13

Comparison of the MultiMap and TSP/CONCORDE Packages for Constructing Radiation Hybrid Maps

Christophe Hitte[1], Travis Lorentzen[2], Richard Guyon[1], Lisa Kim[2], Edouard Cadieu[1], Heidi G. Parker[2], Pascale Quignon[1], Jennifer K. Lowe[2], Boris Gelfenbeyn[2], Catherine Andre[1], Elaine A. Ostrander[2] and Francis Galibert[1]

(1)UMR6061, CNRS. 2 av. Pr. Leon Bernard 35043 Rennes Cedex, FRANCE,

(2)Divisions of Clinical Research and Human Biology, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N., D4-100, Seattle WA 98109-1024 USA.

Corresponding author : Francis Galibert. E-mail : francis.galibert@univ-rennes1.fr

Abstract

RH map construction allows investigators to locate both type I and type II markers on a given genome map. The process is composed of two steps. The first consists of determining the pattern distribution of a set of markers within the different cell lines of an RH panel. This is mainly done by PCR amplification and gel electrophoresis, and results in a series of numbers indicating the presence or the absence of each marker in each cell line. The second step consists of a comparison of these numbers, using various algorithms, to group and then order markers. Because different algorithms may provide (slightly) different orders, we have compared the merits of the MultiMap and TSP/CONCORDE packages using a data set of information currently under analysis for construction of the canine genome RH map.

Introduction.

Whole genome map construction is a two-step process: molecular data generation and the resulting data analysis (McCarthy LC 1996). The latter uses computer programs specifically dedicated to the nature of the map under construction. There are three different types of genome maps: meiotic linkage, RH (Radiation hybrid) and physical maps. They differ, in part, in the type

of markers used to make up the map, the method of genotyping, and the presentation of the results. One of the fundamental differences between meiotic linkage and RH map construction versus physical maps is in assembly methodology. For a physical map, the respective position of two markers A and B is not — or should not be — affected by the addition of new markers to the data set. By comparison, in meiotic linkage and RH map construction, the addition of new markers to an existing data set can, and often does, affect the position of previously mapped markers. This is due to the fact that meitoic linkage and RH maps result from a statistical treatment of experimental data, and thus depend on the analysis program used as well the underlying parameters used in evaluating the data set. As happens frequently, distinct analysis may yield statistically valid, yet distinctly different maps. Even re-computing the same set of data using an identical setting of parameters and the same computer program can produce different versions of a given map (Fig 1).

Radiation Hybrid Mapping.

RH maps result from comparing marker distribution within collections of hybrid cell lines that were previously obtained by fusion of gamma irradiated cells with heterologous carrier cells (Goss and Harris1975; Walter et al 1994, Vignaux et al 1999). Since each viable hybrid contains only a subset, ideally 25% to 35%,of the irradiated genome, markers sharing identical or similar distribution within the RH panel will be identified as being in close physical proximity on the chromosome of interest, while markers with a distinct distribution pattern, are, of necessity, unlinked. During the first step of RH map construction, the presence or absence of each marker to be localized is determined for each cell line of interest by PCR amplification using DNA isolated from each cell line in the RH panel. The resulting data set consists of a series of numbers, – in which '1' indicates the presence of a marker in a specific hybrid-cell line, '0' its absence and '2' an uncertain result. Thus, the distribution of each marker in the panel is characterized by a sequence of 1, 0, 2 called "vectors"(Cox et al 1990), as shown in Fig.2.

During the second step of map construction, marker retention patterns within the panel are compared using different algorithms. This comparison is performed in two phases. In the former, a two-point analysis assigns markers to RH groups that ultimately will correspond to individual chromosomes. In a well developed map there will be only one RH group associated with each chromosomes. The second phase involves determining the markers order within each RH group.

To perform these computations, several computing program packages including RHmap, RHmapper, and MultiMap have been made publically available. (Boehnke et al 1991,Slonim et al 1997, Matise et al 1994).

Interpreting RH maps.

As discussed previously, the end result of RH map construction is a graphical representation of the vector distribution that most closely fits the results of statistical treatment. Unfortunately, for a given set of vectors, there is no unique statistically sound graphical representation. As shown in Fig. 1, the same set of vectors have been successively computed ten times with the MultiMap program (Matise et al 1994), varying only the initial pair of markers that were used. Comparison of the 10 maps shows they are not exactly the same. For instance, the most telomeric marker is marker '8' in six maps, marker '17' in three maps and marker '13' in the last map. Figure 1 also shows that marker '12' is mapped at four different positions 2, 3, 11 and 15. Other discrepancies between the 10 maps can be detected in Fig.1

General Principles of RH map Construction

Two methods, essentially, classified as nonparametric and parametric are widely used in constructing RH maps. Nonparametric methods utilized by program such as RHmap developed by M. Boehnke or program developed by A. Ben-Dor (Boehnke et al 1991, Ben-Dor and Chor 1997) aim to determine the order of markers that minimizes the number of obligate chromosome breaks (OCB). These data are calculated by publically available software based on the retention pattern of each marker. Parametric methods (MultiMap ,RHmapper) (Matise et al 1994, Slonim et al 1997) are based on the comparison of the likelihood of several locus orders. Starting with a pair or triplet of markers, parametric approaches carry out local extension and perform local permutations of consecutive markers to produce the most likely marker order.

RH mapping and the TSP Approach

In 2000 Agarwala et al published an RH computation package using the CONCORDE algorithm that utilizes the TSP or "traveling salesman problem" approach for ordering markers within a specific region (Ben-Dor and Chor 1997, Agarwala et al 2000). In the classic TSP problem, one attempts to determine the shortest route by which a series of cities can be visited,

without ever visiting the same city twice. In the mathematical adaptation of this problem to genome map construction, the cities correspond to the markers and the cost to the distances. The TSP/CONCORDE algorithm systematically computes five independent RH maps. Three are variants of the MLE approach and two of the OCB approach. Agarwala et al. described the TSP/CONCORDE package as an improved option to compute maps resulting in marker orders with higher MLE and lower OCB values. In this particular case, the analysis is insensitive to the initial RH data file and the final map orders are independent of the initial format of the data set (alphabetical order or reverse etc...) as marker order is determined using large neighborhood rearrangements, rather than local permutations. Thus, the work represents a major step forward in RH map building software.

Constructing canine RH maps

We are presently using the TSP/CONCORDE package to order the 3450 markers that make up the most recent version of the whole genome canine RH map (Guyon et al in prep). Figure 3 shows an example of the type of results we have obtained thus far. In contrast to the example presented in Fig 1, the five TSP maps are derived using both principles, i.e. the MLE approach for the first three maps with three independent parameter settings, and the OCB approach with two independent settings to compute the two OCB maps. In its original presentation, the TSP/CONCORDE package (Agarwala et al 2000) presents the results as five independent maps, systematically and automatically generated from the same set of RH data. We developed an additional feature that evaluates the five maps and calculates a consensus map. Our method consists of determining the frequency of the position of a given marker over the five variant maps. When the position of a marker is concordant between the five maps, the placement is considered to have a high confidence level and assigned a support score of 100%. By comparison, any marker displaying a concordant position in only three maps is assigned a 60% confidence level. We then generate a consensus map containing the markers placed at their best position determined by the position frequency calculated among the five TSP maps, as represented in figure 3. Markers with a high confidence support are very likely to be mapped at a robust position, whereas markers present less than three times at the same position in the 5 maps (<60% confidence support), are considered questionable. Occasionally a single marker will be placed at two different positions, revealing a major mapping conflict. Since presentation of the results as a

consensus map is prone to mask regions where a certain level of uncertainty exists we include a graphical representation of the best position data for each marker. This will allow map users to immediately spot region(s) with high statistical support as well as those for which less confidence can be obtained. Interestingly quite often, even if the scrambled region is made of several markers, it can be sub-devised in smaller sub-regions of 2 to 4 markers. (Fig.4 a). At this stage, it is not necessary to account for this slight scrambling; all markers have been typed twice and demonstrate results above a predefined quality threshold. As shown in Fig 4b, by exclusively re-computing the vectors of the 12 markers present between positions 13 and 25 with the TSP/CONCORDE package, more marker placements are now concordant between the five maps.

One final strategy we propose to use for solving construction problems in difficult regions is to repeat the two point analysis using Multimap, but with a higher Lod score than used previously, i.e. 9, 10 or even higher, if the first one was done at Lod 8. When this is done, more than one RH group often results, dividing the chromosome into two to three RH groups. These individual RH groups can often be ordered in a more satisfactorily way with TSP/CONCORDE. Alternatively, —or in addition — a two-point analysis performed with a higher Lod score threshold might eject marker(s) with dubious vector(s), thus facilitating subsequent correct ordering of the novel RH group(s).

Conclusion

It is may be still too early to judge the merits of the TSP/CONCORDE package in RH map building relative to other programs such as MultiMap or RHmapper, which have been extensively used for previous map construction. (Stewart EA et al 1997; Deloukas et al 1998; Priat et al 1998; Mellersh et al 2000; Murphy et al 2000 ; Avner et al 2001 ; Breen et al 2001). However, the advantages we presently perceive manifest themselves at both (i) map construction and (ii) map utilization levels. During map construction, this program acts as an automatic alert highlighting construction problem(s). Such problems can then be solved by regional re-computing and identifying problematic vector(s). Obviously, as shown in Fig 1, several computations of the same vectors can, in principle, be done with another program resulting in delineation of problematic regions. But then this is done using a unique approach and each time with the same parameter setting. In addition such multiple computations are not made automatically and necessitate a program adaptation. At the level of map utilization, graphical

representation of the five maps and display of the name of the markers immediately tell users what confidence they may have in the map and where problems may still exist.

References :

Agarwala R, Applegate DL, Maglott D, Schuler GD, Schaffer AA, 2000. A fast and scalable radiation hybrid map construction and integration strategy. Genome Res 3:350-364.

Avner P, Bruls T, Poras I, Eley L, Gas S, Ruiz P, Wiles MV, Sousa-Nunes R, Kettleborough R, Rana A, Morissette J, Bentley L, Goldsworthy M, Haynes A, Herbert E, Southam L, Lehrach H, Weissenbach J, Manenti G, Rodriguez-Tome P, Beddington R, Dunwoodie S, Cox RD, 2001. A radiation hybrid transcript map of the mouse genome. Nat Genet. 29:194-200

Ben-Dor A and Chor B, 1997. On constructing radiation hybrid maps. J Comput Biol. Winter; 4:517-533

Breen, M., Jouquand, S., Renier, C., Mellersh, C. S., Hitte, C., Holmes, N. G., Chéron, A., Suter, N., Vignaux, F., Bristow, A. E., Priat C., McCann E., Andre C., Boundy S., Gitsham P., Thomas R., Bridge W., Spriggs H.F, Ryder E.J., Curson A., Sampson J., Ostrander E.A., Binns M., and Galibert F, 2001. Chromosome-specific single-locus fish probes allow anchorage of an 1800-marker integrated radiation-hybrid/linkage map of the domestic dog genome to all chromosomes. Genome Res. 11:1784-1795.

Boehnke M, Lange K and Cox DR, 1991. Statistical methods for multipoint radiation hybrid mapping. Am J Hum Genet 49:1174-1188.

Cox DR, Burmeister M, Price ER, Kim S and Myers RM, 1990. Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. Science 250:245-250.

Deloukas P, Schuler GD, Gyapay G, Beasley EM, Soderlund C, Rodriguez-Tome P, Hui L, Matise TC, McKusick KB, Beckmann JS, Bentolila S, Bihoreau M, Birren BB, Browne J, Butler

A, Castle AB, Chiannilkulchai N, Clee C, Day PJ, Dehejia A, Dibling T, Drouot N, Duprat S, Fizames C, Bentley DR, et al. 1998. A physical map of 30,000 human genes. Science 282(5389):744-746.

Goss SJ and Harris H, 1975. New method for mapping genes in human chromosomes. Nature 255:680-684.

Matise TC, Perlin M, Chakravarti A, 1994. Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linkage map Nat Genet. 4:384-90.

McCarthy LC, 1996. Whole genome radiation hybrid mapping. TIG 12:491-493.

Mellersh CS, Hitte C, Richman M, Vignaux F, Priat C, Jouquand S, Werner P, Andre C, DeRose S, Patterson DF, Ostrander EA, Galibert F, 2000. An integrated linkage-radiation hybrid map of the canine genome. Mamm Genome 11(2):120-130.

Murphy WJ, Sun S, Chen Z, Yuhki N, Hirschmann D, Menotti-Raymond M, O'Brien SJ, 2000. A radiation hybrid map of the cat genome: implications for comparative mapping. Genome Res 5:691-702.

Priat C, Hitte C, Vignaux F, Renier C, Jiang Z, Jouquand S, Cheron A, Andre C, Galibert F, 1998. A whole-genome radiation hybrid map of the dog genome. Genomics 54(3):361-378

Slonim D, Kruglyak L, Stein L, Lander E, 1997. Building human genome maps with radiation hybrids. J Comput Biol Winter;4:487-504.

Stewart EA, McKusick KB, Aggarwal A, Bajorek E, Brady S, et al, 1997. An STS-based radiation hybrid map of the human genome. Genome Res 7:422-433.

Vignaux, F., Hitte, C., Priat, C., Chuat, J.C., André, C., and Galibert, F, 1999. Construction and optimization of a dog whole-genome radiation hybrid panel. Mamm. Genome 10: 888-894.

Walter MA, Spillett DJ, Thomas P, Weissenbach J and Goodfellow PN, 1994. A method for constructing radiation hybrid maps of whole genomes. Nat Genet 7:22-28.
TSP/CONCORDE package : (ftp://ftp.ncbi.nih.gov/pub/agarwala/rhmapping/rh_tsp_map.tar.gz)

Figure Legends :

Figure 1 :

Vectors, corresponding to a subset of markers located on a canine chromosome have successively been computed ten times with the MultiMap program, changing only the initial pair of markers used for the computation. Comparison of these 10 maps shows that although a given marker is predominantly present at each position (i.e.marker 8 is present 7 times out of 10 in position 1, marker 17 is present 3 time and marker 13 one time), no position is occupied in all 10 maps by the same marker.

Figure 2 :

Example of vector suites defining the pattern distribution of markers within a RH panel. Presence of a marker in a specific cell line is indicated by 1, its absence by 0, and uncertain results are noted by a 2.

Figure 3 :

The vectors grouped to a given chromosome by 2 point analysis performed with MultiMap were then ordered with TSP/CONCORDE automatically delivering 5 maps. Results of the comparison of these five maps is enlighted by horizontal bars. When the same marker is present in the five maps at a given position the horizontal bars have a maximum length and correspond to mapping positions reaching high confidence (i.e. box 1). Map positions occupied by two or more markers are characterized by shorter horizontal bars (i.e. boxes 2 and 3). Although as in box 2 a given marker can occupied three different positions extending the size of the scrambled region, in box 3 only two adjacent markers exchange their positions limiting the zone of uncertainty. Such results probably reflect the overall resolution of the 5000 rad panel used in these experiments.

Figure 4 :

a)- TSP/CONCORDE computation results presented with horizontal bars to indicate the level of agreement between the different maps. The box corresponds to a region of 12 markers where discrepancies between the five maps are observed. Nevertheless, this region can be subdevised in 4 definable sub-regions, indicated by vertical bars, limiting the extent of the discrepancies.
b)- The results obtained by re-computing the 12 boxed markers are shown. Note that re-computation of a limited number of vectors can result in a higher level of confidence within a regional map. .

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of the American Kennel Club Canine Health Foundation, the Burroughs Wellcome Fund, the U.S. Army, grant DAAD19-01-1-0658 and NIH, grant 1R01CA092167 to E.A.O and F.G. R.G. is supported by a fellowship from « La Region de Bretagne », L.K. is supported by a fellowship from Nestle Purina and H.G.P by PHS National Research Grant T32 GM07270.

Figure 1 : The ordering step with the MutiMap package.

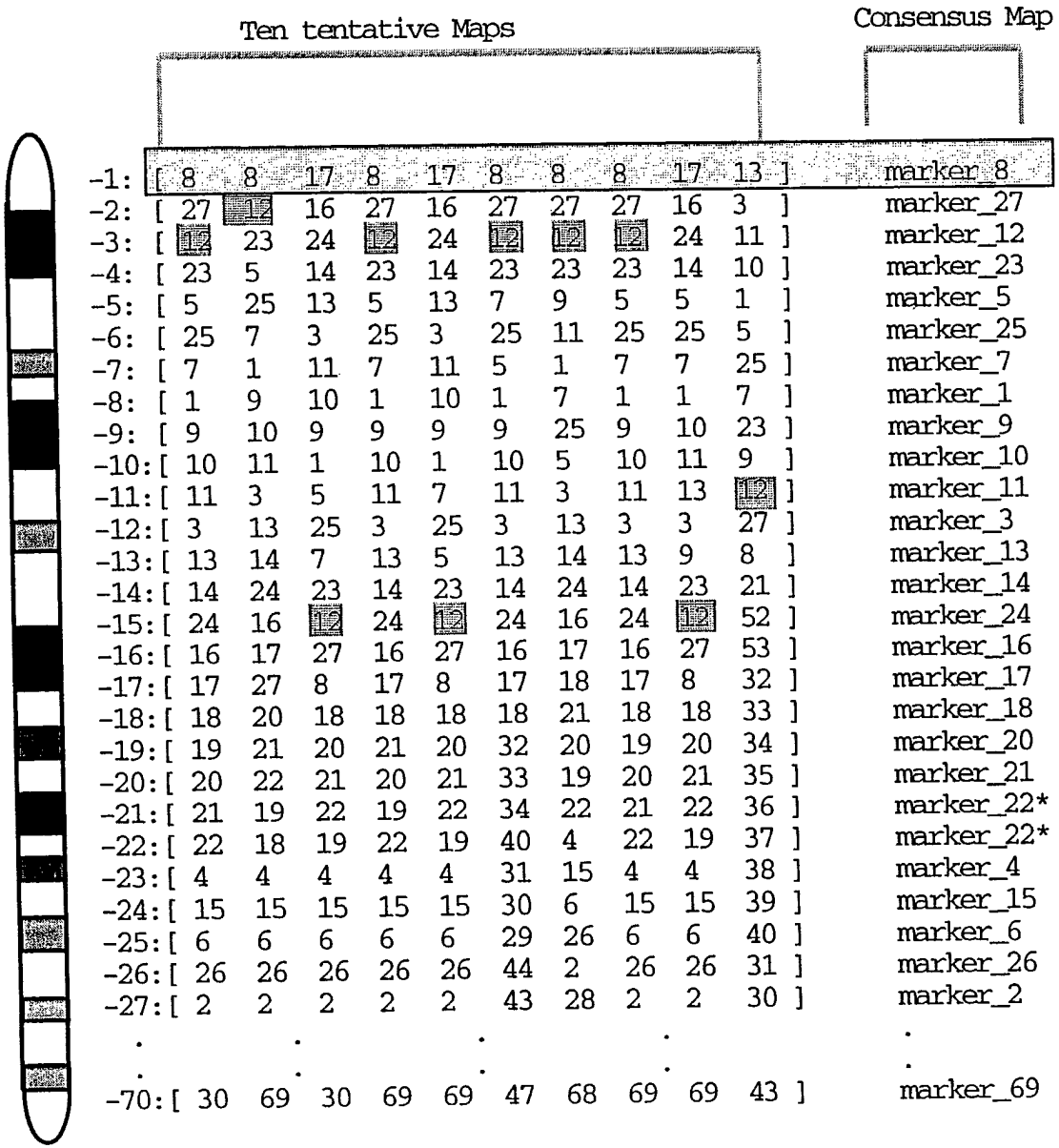


Figure 2 : RH vectors flat file.

Markers Vectors

```
Mker_1  00000000000000000111000001000010000000000000000010001000000010001000000000000000000000010000000000000000000000200000
Mker_2  10000010010000000110000010000000000000000001001000000000001001000100010120000010010000000100011001010101001000000001021
Mker_3  10000000012000000000001000000001000000000100000000010000000000000000010000000001010000010000000000000000000000010000
Mker_4  01000000000000000011000000100000000000001000000000000010001010100000000000000000000010000000000000000000000000100001
Mker_5  00000010000000000100000000000001000000000000000000010101000100000010022000000000000010000000000000010010000000001101
Mker_6  00000000010000000011000000000100100000000100000000010100000002000000000000000000101000000000000010000000000000100001
Mker_7  10000000100000000010000000000011000000000000000000000000000001000010020120000000000000010100000000000000000000000000
Mker_8  00000101000000000010000000000000111000000000000000000010000000010101001000000000000000000000000000000000000000001001
Mker_9  00000201000000000000000000000000000000000000000000000000000001000000000000000000001000010000000000001000000000000001
Mker_10 00000200000000010000000000000000000000000000000000000001000000000000010000000000000000000010000000001000000000000000
Mker_11 00000020000000010000000000000000100000000000000000000000000000000001000000000000000100001000010100000001000000000000
Mker_12 00000021000000000100000001000100000001000000000000000000000100000000100000000010000000000100000000010100000000001000
Mker_13 10000001000000000001010000010010000000000000011010000100000001001000000000001000000100000000011100000020000001101010000
Mker_14 00000000001001000010010001000000100000100000000010000010000100000001101100001000110000001000000000000000000000000001
Mker_15 01000201000000100110000200010011100000000000000100000101000001000100000000000000000000010100000010000000000000000001
Mker_16 00010210000000000000100100000000000000000000010000000001000000010100001010011210000100000000000000000000010000001000000
Mker_17 01000021010100000110000010000000100000000001000000000011000001010101010000000100000000000100000000100100000000000110000
Mker_18 00000020000100000100000010000000000000000000000000000001000001010101010000000000000000000000000000000010000000000110000
Mker_19 01000021010000000110000000000000100000000010000000001000000100000000000000010000000000100000000100100000000000010000
Mker_20 11000010110000000110000100000000000000001001100000000010010000000000000010000011010010001110010000010100000001101000001
```

Figure 3 : A method for evaluating map order with the TSP/CONCORDE approach.

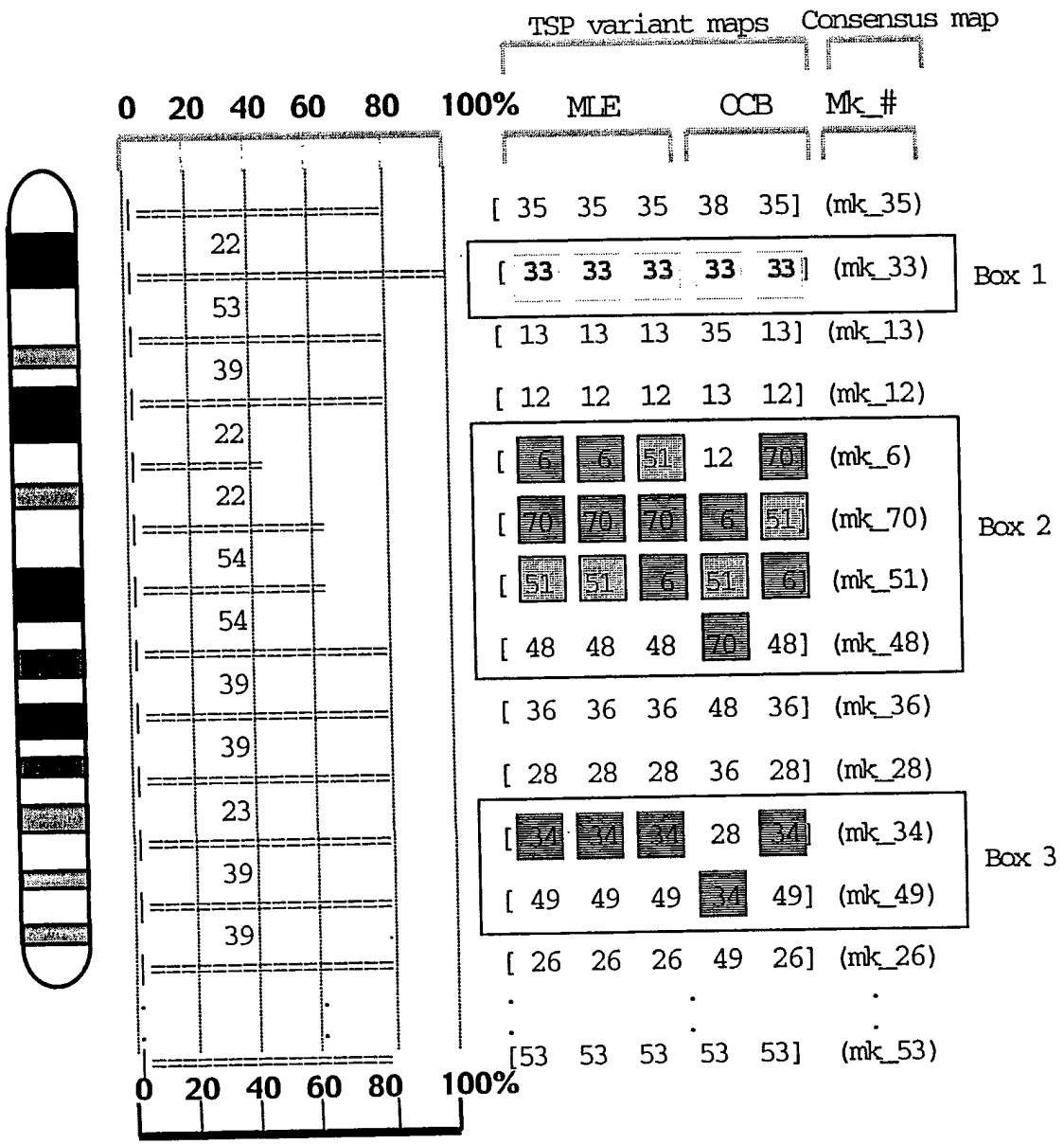
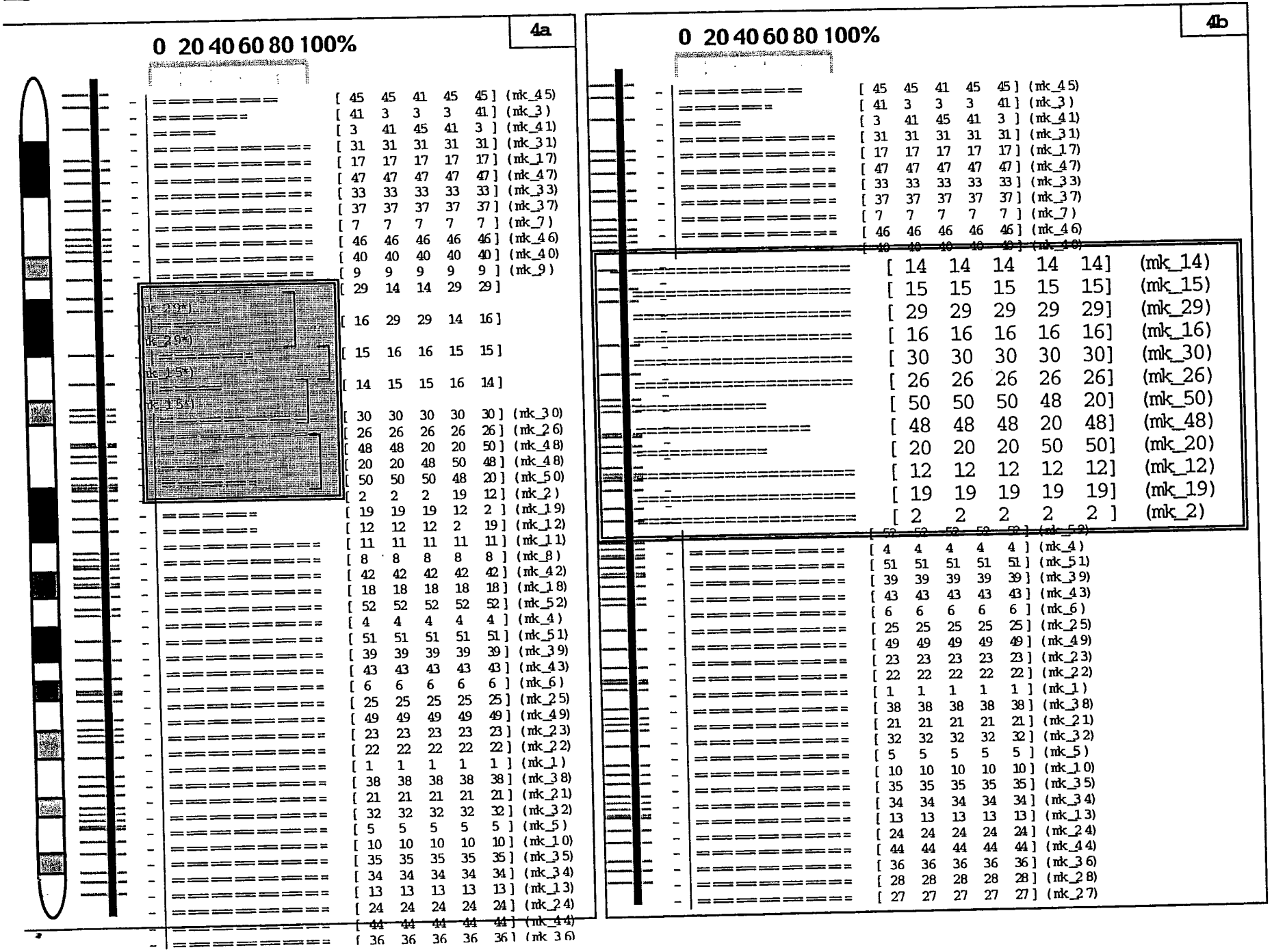


Figure 4 : The TSP approach for RH map construction



Comparative Mapping of Human Chromosome 1p and the Canine Genome

(Key Words: canine, 1p, comparative map, radiation hybrid)

Richard Guyon[a], Ewen F. Kirkness[b], Travis D. Lorentzen[c], Christophe Hitte[a], Kenine E. Comstock[c], Pascale Quignon[a], Thomas Derrien[a], Catherine André[a], Claire M. Fraser[b], Francis Galibert[a] and Elaine A. Ostrander[c]

[a]UMR 6061 CNRS, Génétique et Développement, Faculté de Médecine, 35043 Rennes Cédex, France

[b]The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD, USA 20850

[c]Clinical and Human Biology Divisions, Fred Hutchinson Cancer Research Center, P.O. Box 19024, D4-100, Seattle, WA, USA 98109-1024

Pre Publication Correspondence: Elaine A. Ostrander, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N., D4-100, PO Box 19024 Seattle WA, 98109-1024. Email: eostrand@fhcrc.org; Phone: 206-667-6979; FAX: 206-667-6396

Running Title: Comparative Mapping of Human chromosome 1p in the Dog

Number of Pages 21; Figures 1; Tables 1; Words 5,466

ABSTRACT

The ability to test the association between phenotype and genotype within biological systems of interest is limited by the degree to which the genome map of any model system can be rigorously aligned with the reference human and mouse maps. While extensive reciprocal chromosome paint studies have outlined the general evolutionary relationships between the chromosomes of dog and other mammals, details of the conserved synteny that exist between the human and dog genomes is still lacking. Towards this end, we have tested the hypothesis that 1x sequence coverage of the canine genome is sufficient to allow identification and mapping of canine orthologs for most human genes. In the following study, we define the evolutionary relationships between the canine genome and human chromosome 1p (HSA1p). The definition and mapping of 120 novel canine genes, orthologous to HSA1p genes, allowed identification of seven conserved segments within five chromosomal regions (Canis Familiaris chromosomes (CFA) 2, 5, 6, 15 and 17). The resolution of conserved segments, and the establishment of gene order within these segments, facilitated construction of a detailed comparative map between human and dog in this region of interest. The study presented here, therefore, illustrates the power of combining 1x shotgun sequence data and a one megabase resolution radiation hybrid (RH) map for building a comparative map with the human sequence. The net result is a unified resource suitable for studies aimed at positional cloning of mapped loci, candidate gene assessment, and evolutionary analyses.

INTRODUCTION

The ability to move from linked markers to candidate genes within any model genome mapping system is limited at one level by the number of informative markers on the map, and at another by the degree to which the genome map of interest can be aligned with the human and mouse reference maps. In the case of the canine genome, reciprocal chromosome painting has enabled investigators to broadly establish the evolutionary relationship between canine chromosomes and cytogenetic bands defining human chromosome arms (Breen 1999; Yang 1999). These data suggest the existence of 68-73 conserved regions (Breen 1999; Sargan 2000; Yang 1999) while radiation hybrid data suggest a total of 76 conserved segments between human and dog (Guyon 2003). Radiation hybrid data demonstrate that several canine chromosomes, such as *Canis familiaris* chromosomes (CFA) 8, 12, 22-24, and most of the smaller canine chromosomes are apparently comprised of a single continuous section of the human genome; others such as CFA1 through CFA7 retain two to four portions of several distinct human chromosomes, and still others, such as CFA15, correspond to as many as five HSA fragments (Guyon 2003). This fact, combined with the nearly 1600 microsatellite markers now ordered on the canine radiation hybrid (RH) map (Guyon 2003; Parker 2001), ensures that genome-wide scans on informative canine families can be carried out with relative ease, and the corresponding chromosome arm in the human genome can be quickly and correctly identified.

However, within the large syntenic regions that define each canine chromosome, there is a paucity of mapped genes that severely limits the ability to move from a general region of interest to selection of specific candidate genes. Indeed, only 900 canine-specific genes have been placed on the most recent version of the canine radiation hybrid (RH) map (Guyon

2003), and still fewer on the meiotic linkage map (Parker 2001). Overall, the distribution of gene-based markers averages only one per 3 Mb. While these data support the hypothesis that blocks of several megabases are well conserved throughout the canine genome, the number of mapped genes within any single block is insufficient for assigning breakpoints of conserved synteny. This limits the degree to which initial findings of linkage in canine families can be followed by successful positional cloning efforts, and reduces the utility of the human genome sequence for tackling problems of interest in other mammalian systems. Thus, it remains a priority of the canine genome mapping community to add more gene-based markers to the canine map.

In this study, we have tested the hypothesis that 1x sequence coverage of the canine genome is sufficient to permit identification and mapping of the canine orthologs of most human genes. Towards this aim, we focused on the human chromosome 1p arm (HSA1p), which is known to contain several disease-associated genes of interest. The 1x sequence was used to identify canine orthologues of 158 genes from HSA1p, and RH mapping of 120 of them allowed production of a dense comparative map between human and dog in this region of interest. Human HSA1p corresponds to seven conserved segments within five chromosomal regions (CFA 2, 5, 6, 15 and 17) with gene orders and limits well defined. Our efforts to map a total of 161 well-distributed genes of HSA1p demonstrate the feasibility of using a 1x sequencing resource to derive dense comparative maps. We suggest that, for many additional genomes, this will be a powerful and economical approach for characterizing genome structure and evolutionary relationships.

RESULTS

Starting from a set of 187 HSA1p genes, fragments of 158 putative orthologues were retrieved from the canine genomic sequence data. For 144 genes, canine specific primers

were designed and 126 were successfully typed on the RHDF5000-2 canine radiation hybrid panel (Vignaux 1999). RH data from the markers were computed with the latest 3270-marker RH map (Guyon 2003) using the MultiMap and Traveling Salesman (TSP)/CONCORDE softwares (Matise 1994; Agarwala 2000). Of these 126 gene markers, 120 could be incorporated in five of the 38 canine autosomes and six markers remained unlinked (Table A on website). The five chromosomes that are shown by this analysis to correspond to HSA1p are: CFA 2, 5, 6, 15 and 17. Markers were then ordered on each chromosome using the TSP/CONCORDE software. These 120 gene markers were added to the 41 previously mapped (Guyon 2003), bringing the total number of canine gene markers to 161 in these five chromosomal regions (Table1). However, for four of those previously mapped genes, no human counterparts could be found on the NCBI Build 31 database of the human sequence, thus 157 gene markers constitute informative anchor sites between the human and dog genomes. The current comparative map between HSA1p and the canine orthologous regions is shown on Figure 1.

The total number of canine gene-based markers assigned to each of the five chromosomal regions orthologous to HSA1p ranges from 52 on CFA6 to 14 on CFA17, while the total number of canine markers ranges from 78 on CFA6 to 18 on CFA17 (Table 1). On CFA6, however, the 78 markers are mapped to 41 unique positions, while on CFA17 the 18 markers are mapped to 10 unique positions. The increase in the number of markers that are co-positioned when the total number of markers mapped to some regions increase reflects the limited resolution of the RHDF5000-2 panel. As indicated in previous studies (Breen 2001; Guyon 2003), the panel resolution has been estimated to be approximately 600 kb, therefore markers present in an interval of one Mb cannot be accurately ordered with respect to their immediate neighbors (Priat 1998).

As shown in Figure 1, we noted different degrees of conserved homologous gene association as defined by the First International Workshop on Comparative Genome Organization (Andersson 1996). Conserved synteny, defined as the association of two or more homologous genes in two species, regardless of gene order or inter-spacing of noncontiguous segments, has been observed for CFA 2, 5, 6, 15 and 17. Conserved segments are defined as the syntenic association of two or more homologous and contiguous genes not interrupted by different chromosome segments in either species. The HSA1p orthologous region of CFA 15 is split in two conserved segments by an asyntenic fragment. Conversely, the CFA 5 orthologous region of HSA1p is made of two conserved segments separated by an asyntenic fragment. The HSA1p orthologous regions of CFA 2, 6, and 17 are made of only one conserved segment, with a part of the CFA17 conserved segment being inverted. Conserved order is defined as the demonstration that three or more homologous genes lie on one chromosome in the same order in two species. In this study, conserved orders were observed in the conserved segments of CFA 5, 6 and 15. Indeed, a detailed screen for gene orders allowed us to split both CFA 2 (CS II) and CFA17 (CS VII) conserved segments into two distinct blocks where gene order is head-to-head.

A detailed analysis of CFA 5 and 15 shows that the syntenic association of homologous genes is contiguous only in one of the two species (Figure 1). For CFA5, CS I and CS V, separated by 42.6 Mb in the human sequence appear as a single contiguous block in dog. CS I is identified by 13 anchor sites and spans 8.8 Mb on HSA1p. CS V is identified by 24 anchor sites and spans 13.8 Mb (Table 1). The gene order inside each block is conserved between human and dog, but CSI is inverted relative to HSA1p. For CFA15, the HSA1p orthologous region is split into two conserved segments of 14 and 10 anchor sites by an asyntenic region 328 TSP units long, and composed of three markers orthologous to a 0.2 Mb interval of HSA16. The gene orders in these two conserved segments are in the same

orientation but their relative positions reveal a transposition event in the dog genome (Figure 1).

CS VII on CFA17 is composed of two blocks of 5 and 14 anchor sites respectively, the second one overlapping the centromeric region of HSA1 and being in inverted orientation, includes 6 genes of HSA1q. Although gene orders within each of these blocks cannot be assessed with certainty because of the RH panel resolution limit, this organization reveals an inversion event including the centromeric region of HSA1 or in the CFA17 orthologous region. The whole conserved segment between CFA17 and HSA1 represents 1067 TSP units in the canine map. In human, it includes the 21 Mb of the centromere of HSA1 (HSA1p11.1-q12) and spans 39.8 Mb.

In contrast to the above, we see little variation associated with CFA2 and CFA6, where 35 and 54 anchor sites have been mapped, respectively. The association of homologous genes in human and dog is contiguous and in accordance with the definition of conserved segments. Finally, we note that the gene order between HSA1p and CFA6 (CS VI) is entirely conserved. However, CS II is subdivided in two blocks of conserved order, harboring 23 and seven anchor sites, respectively, highlighting an additional rearrangement in one of the two species.

DISCUSSION

Using a simple statistical model, it can be estimated that the 1.2x coverage of the dog genome will provide 70% of the genome sequence, with an average gap length of ~480 bases (Lander and Waterman 1988). It can also be estimated that sequences containing 100 bp of an exon (or exon fragment) will be sufficient to identify dog orthologues of most human genes. The probability of a specific 100-base fragment of the genome occurring entirely within a single sequence read of 576 bases is only 0.58. However, most human genes appear

to be composed of at least 4 exons (Venter 2001), and given the known similarity in gene structure between humans and canids (for example: (Credille 2001; Haworth 2001; Szabo 1996), the same is likely to be true for dog. Consequently, the probability that at least one 100-base exon fragment from a gene is contained within the genomic sequence data can be estimated as >0.95. It has not yet been determined what proportion of human genes for which 1:1 orthology can be detected in the dog genome. For mouse, this value has been estimated to be approximately 80% (Okazaki 2002; Waterston 2002). If dogs and humans share a similar number of orthologous genes, we can estimate that the dog genomic sequence data will yield at least one orthologous exon fragment for ~80% of human genes. In this study, fragments of putative dog orthologues were identified for 158 of 187 (84%) of selected human genes. Recently, a more comprehensive analysis has indicated that 79% of all annotated human genes (and 96% of those that have detectable orthologues in mouse) are represented by orthologous dog sequences in the 1.2x coverage (Kirkness et al., unpublished data).

If the primary objective of a sequencing project is to generate gene-based markers for RH-mapping,1x sequence coverage of a genome offers several advantages over large collections of ESTs. Unlike cDNA libraries, the representation of genes is unaffected by cellular expression levels, and identification of orthologous exons is not biased by the length of 3'-untranslated mRNA. In addition, the low but significant conservation of intronic sequences between species is useful for distinguishing between paralogous sequences that share substantial sequence identity within exons.

The most recent iteration of the canine RH map (Guyon 2003) featured 870 markers for which orthologous sequences have been identified on the human genome. Although the HSA1p orthologous regions were shown to correspond to five canine chromosomes, CFA 2, 5, 6, 15 and 17 in the previous RH map, by increasing more than 4-fold the number of

markers, this work clearly delineates the gene order and breakpoints for seven conserved segments. The largest increase in resolution is in the HSA1p orthologous region of CFA5 (CS I and V) which now contains 34 genes compared to four in the previous version of the canine map.

This comparative map allows us to characterize more precisely the conserved segments orthologous to HSA1p that were previously identified by reciprocal chromosome painting studies (Breen 1999; Sargan 2000; Yang 1999) on CFA5, CFA15 and CFA17 (Figure 1). On CFA5, while contiguous in dog, two conserved segments (I and V) are split in human. On CFA15, the region is split by a novel asyntenic fragment of HSA16 as previously shown by Guyon (Guyon 2003), leaving two conserved segments (III and IV) harboring inverted positions in human versus dog. Finally, on CFA17, the region is split into blocks but constitutes a unique conserved segment. The two remaining canine chromosomal regions (CFA2 and CFA6) each constitute a unique conserved segment.

The seven regions span 115 Mb of the 123 Mb HSA1p arm, indicating that for roughly 8 Mb of HSA1p, no canine counterparts have yet been mapped. Between those conserved segments, six regions that contain the breakpoints of interest range in length from 0.3 to 3.8 Mb, and represent a total of 7.3 Mb. Together, those intervals contain 113 human genes (http://www.ncbi.nih.nlm.gov/mapview) ranging from 58 genes in the 3.8 Mb region between CS II and III to 6 genes in the 0.4 Mb region between CSIII and IV. In addition, 42 genes are present in the 1 Mb most telomeric region of HSA1p above CS I. Despite the high density of anchor sites along HSA1p (1/700 kb), eight intervals greater than two megabases with no mapped genes in dog still remain inside conserved segments. The two largest span 7.3 Mb in CS VII and 6.5 Mb in CS VI and contain 40 and 50 genes, respectively. The other six intervals spanning less than 3 Mb contain from 3 to 35 genes. These intervals are likely to contain additional conserved segments that will be resolved by RH mapping additional genes

retrieved from the 1x canine sequence. Additional sequencing can be done to more clearly delineate the breakpoints.

This comparative map has allowed us to compare gene orders in human and dog, and to comment on possible intra-chromosomal rearrangements. In five of the seven conserved segments, the gene orders are strictly conserved, while CS II on CFA2 contains a small inverted segment. Despite the fact that gene order cannot be precisely assessed in CS VII, the two blocks probably harbor inverted gene order as a consequence of the chromosomal inversion that brought HSA1q orthologous genes between HSA1p orthologous blocks.

On the current canine RH map, some local discrepancies leading to an artifactual inversion of local orders are observed. This is likely due to the resolution limit of the RHDF5000-2 panel, estimated to be about 600 kb (Vignaux 1999). A related problem, the high number of co-localized anchor sites, especially on CFA6 (CS VI), highlights the saturation of the canine HSA1p orthologous map in discrete regions. The use of a higher resolution canine RH panel would allow us to circumvent both problems. Some local discrepancies in the comparative map are, however, likely due to slight distortions in the human sequence assembly, typically observed when updating the human localization of anchor sites from one NCBI Build to the next. Indeed, according to NCBI build 31, the HSA1p arm is still composed of at least 56 contigs, separated by gaps of unknown size and sequence.

In order to date evolutionary breakpoints between human and dog identified in HSA1p and to establish in which lineages such events happened, the comparison of the conservation between HSA1p and various mammals is very instructive. The ancestral genome of primates and carnivores was likely a low-numbered, largely metacentric genome that evolved at a slow rate to human (11 steps), cat (6 steps), mink (10 steps), and seal (8 steps) (O'Brien 1999). Chromosome rearrangements can be used as characters for phylogenetic reconstruction

following the principle of outgroup comparison (Yang 2000). The HSA1p region appears to be entirely syntenic between human and cat (Murphy 2000). This indicates that the split into five chromosomal segments in dog occurred in the Canoidea lineage following the Canoidea and Feloidea radiation, some 60 million years ago (Wayne 1993). Yang and colleagues (Yang 1999) showed by reciprocal chromosome painting that HSA1p is also split in five chromosomal segments in the red fox, indicating that these evolutionary events occurred before the dog and red fox divergence, some ten million years ago (Wayne 1993). This time estimate could be refined by the comparison of genomic rearrangements between human and other canoidea superfamily members, provided an appropriate comparative map with human is well established.

While the mammal radiations generally display a slow rate of chromosome exchange, approximately 1-2 exchanges per 10 million years, certain lineages show a more rapid pattern of chromosome change. Consider for example the primate lineage, in which the genome is mostly conserved between human, chimpanzee and macaque, while it is dramatically shuffled in the gibbon lineage (O'Brien 1999; O'Brien and Stanyon 1999). Similarly, in the carnivore lineage, the dog, as well as other canids, have appreciably rearranged genomes relative to the ancestral carnivore organization, indicating a high rate of chromosome exchange (Wayne 1987; Wayne 1987; Yang 1999). Although only HSA1p orthologous regions are considered here, this study suggests similar findings.

In this comparative map, an HSA16 orthologous region is found contiguous or within HSA1p orthologous regions in four out of five instances. The HSA16 conserved segments are found contiguous to HSA1p in CFA2, 5 and 6 while a small conserved segment is found inside the HSA1p region in CFA15. In Carnivora, this association is not found in cat, arguably since its genome is less rearranged and very close to human in this region (Murphy

2000; Yang 2000). We have no explanation for this association; it may be a consequence of poorly understood evolutionary forces, or merely a coincidence.

Detailed comparative maps between closely and distantly related species are of great interest in understanding the evolutionary relationships between species, families and orders. The study presented here illustrates the joint utility of the 1x shotgun sequence approach and a relatively dense RH map for building a comparative map with the human genome. The net result is a unified resource that can facilitate studies aimed at genetic mapping, positional cloning of mapped loci, and evolutionary studies of species of interest.

METHODS

Selection of Orthologs Derived from Canine 1x Sequence

Sequence from the canine genome was derived as follows: Genomic DNA from a male Standard Poodle was used to prepare plasmid libraries of small- and medium-sized inserts (~2 kb and ~10 kb respectively). End-sequencing of clones from each library was conducted at Celera Genomics as described previously (Venter 2001), and yielded 3.42 million reads (86.7% paired) from 2 kb clones, and 2.81 million reads (86.4% paired) from 10 kb clones. Read quality was evaluated in 50-bp windows using Paracel's TraceTuner, with each read trimmed to include only those consecutive 50-bp segments with a minimum mean accuracy of 97%. End windows (both ends of the trace) of 1, 5, 10, 25, and 50 bases were trimmed to a mean accuracy of 98%. Every read was checked further for vector and contaminant matches of 50 bases or more. The finished sequence data consists of 6.22 million reads (mean read length, 576 bases), representing approximately 1.2x coverage of the 3 Gb haploid canine genome (Vinogradov 1998).

For 187 genes known to span HSA1p, the associated peptide sequence was searched against the complete collection of dog reads using tblastn. For each peptide, all homologous

dog reads that were identified by the blast searches were assembled at high stringency (99% nucleotide identity) using TIGR Assembler (http://www.tigr.org/softlab/assembler/). Each assembly, or unassembled read, was then searched back against the Ensembl (release 1.1) collection of confirmed cDNAs and peptides (using blastn and blastx respectively). If the highest scoring hits (for both the DNA- and protein-sequence searches) were to the gene that was used originally for searching, the assembly was considered a fragment of a putative orthologue. The coordinates of each human gene on HSA1p were obtained from NCBI build 31 of the human genome (http://genome.ucsc.edu).

Radiation Hybrid Mapping

Genes were mapped on the 118 cell lines of the RHDF5000-2 panel previously described (Vignaux 1999). In brief, PCR primers were selected for mapping using a standard selection program i.e. Primer3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). Whenever possible, both primers were selected in the two introns flanking the annotated exon sequence. Alternatively, to better ensure amplification of the correct gene, in some cases one primer was selected from a flanking intron and the other from a corresponding exon. Primers were preferentially selected to be 25 bp in length and to work under a single optimal set of PCR conditions (salt, Tm, Mg^{+2}, etc.) generating PCR products of 200-250 bp.

Typing of markers was done using existing infrastructure described previously (Breen 2001; Guyon 2003; Mellersh 2000; Priat 1998). In brief, all reactions are done using a 96-well or 384-well format in a volume of 10-15 µl. An initial screen using 50 ng dog DNA, 50 ng hamster DNA, and a 1:3 mix of dog/hamster DNA (50 ng) is used to select primers suitable to be placed across the entire panel. PCR reactions were done with 50 ng of RH

DNA and products were resolved on 1.8% or 2% agarose gels, electrophoresed for 30 minutes as described previously (Breen 2001; Mellersh 2000; Priat 1998). Bands were viewed under UV light after ethidium bromide staining, and an image was recorded.

All markers were typed in duplicate and were considered consistent when the two vectors for each marker had a discrepancy value <16%, calculated for each marker based on its retention value within the panel. This threshold of 16% was determined to correspond to a distance lower than the resolution limit of the RHDF5000-2 panel (600 kb). Details and PCR conditions for all markers are available in table A at:

http://www-recomgen.univ-rennes1.fr/doggy.html

http://www.fhcrc.org/science/dog_genome/dog.html

Analysis and Map Construction

Novel markers were incorporated into the latest 3270 marker RH data set (Guyon 2003). The corresponding RH groups were computed by pairwise calculations using the MultiMap software (Matise 1994) at a Lod threshold ≥ 8.0, thus allowing HSA1p orthologous canine gene markers to be assigned to specific chromosomes. In order to refine the region of interest containing orthologous HSA1p genes, the relevant chromosomes were split into smaller RH groups using the MultiMap algorithm and a Lod threshold of >9.0. Contiguous groups of the same chromosome origin were computed together. RH groups containing at least one HSA1p orthologous marker were then ordered using the TSP approach as specified by the CONCORDE computer package (http://www.math.princeton.edu/tsp/concorde.html) (Agarwala 2000). TSP/CONCORDE computes five independent RH maps and the resulting maps were subsequently evaluated to produce a consensus map using a method developed by us (Hitte 2003). Inter-marker

distances were determined with the rh_tsp_map1.0 version of TSP/CONCORDE which produces map positions in arbitrary TSP units.

ACKNOWLEDGEMENTS

We acknowledge the American Kennel Club Canine Health Foundation, U.S. Army Grant DAAD19-01-1-0658 (E.A.O. and F.G.) and NIH R01CA-92167 (E.A.O, E.K. and F.G.) In addition, E.A.O is supported by K05 CA90754 and the recipeint of a Burroughs Wellcome Award in Functional Genomics. R.G. is partly supported by an AKC and CNRS fellowship, and P.Q. by a Conseil Regional de Bretagne fellowship.

REFERENCES

Agarwala, R., Applegate, D.L., D. Maglott, G.D. Schuler and Schaffer, A.A. 2000. A fast and scalable radiation hybrid map construction and integration strategy. *Genome Res* **10**: 350-364.

Andersson, L., Archibald, A., Ashburner, M., Audun, S., Barendse,W., Bitgood, J., Bottema, C., Broad,T., Brown, S., and Burt, D. et al. 1996. Comparative genome organization of vertebrates. The First International Workshop on Comparative Genome Organization. *Mamm Genome* **7**: 717-734.

Breen, M., Jouquand, S., Renier, C., Mellersh, C.S., Hitte, C., Holmes, N.G., Cheron, A., Suter, N., Vignaux, F., and Bristow, A.E. et al. 2001. Chromosome-specific single-locus FISH probes allow anchorage of an 1800-marker integrated radiation-hybrid/linkage map of the domestic dog genome to all chromosomes. *Genome Res* **11**: 1784-1795.

Breen, M., Langford, C.F., Carter, N.P., Holmes, N.G., Dickens, H.F., Thomas, R., Suter, N., Ryder, E.J., Pope, M., and Binns, M.M. 1999. FISH mapping and identification of canine chromosomes. *J Hered* **90**: 27-30.

Breen, M., Thomas, R., Binns, M.M., Carter, N.P., and Langford, C.F. 1999. Reciprocal chromosome painting reveals detailed regions of conserved synteny between the karyotypes of the domestic dog (Canis familiaris) and human. *Genomics* **61**: 145-155.

Credille, K.M., Venta, P.J., Breen, M., Lowe, J.K., Murphy, K.E., Ostrander, E.A., Galibert, F., and Dunstan, R.W. 2001. DNA sequence and physical mapping of the canine transglutaminase 1 gene. *Cytogenet Cell Genet* **93**: 73-76.

Guyon, R., Lorentzen,T.D., Hitte, C., Kim, L., Cadieu, E., Parker, H.G., Quignon, P., Lowe, J.K., Renier, C., and Gelfenbeyn, B. et al.. 2003. A 1 Mb resolution radiation hybrid map of the canine genome. *Proc. Natl. Acad. Sci. U.S.A.* In Press.

Haworth, K.E., Islam, I., Breen, M., Putt, W., Makrinou, E., Binns, M., Hopkinson, D., and Edwards, Y. 2001. Canine TCOF1; cloning, chromosome assignment and genetic analysis in dogs with different head types. *Mamm Genome* **12**: 622-629.

Hitte, C., Lorentzen, T., Guyon, R., Kim, L., Cadieu, E., Parker, H., Quignon, P., Lowe, J., Gelfenbeyn, B., Andre, C., Ostrander, E.A., and Galibert, F. 2003. Comparison of the MultiMap and TSP/CONCORDE packages for constructing radiation hybrid maps. *J. Heredity* In Press.

Lander, E. and Waterman, M. 1988. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* **2**: 231- 239.

Matise, T.C., Perlin, M., and Chakravarti, A. 1994. Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linkage map. *Nature Genet.* **6**: 384-390.

Mellersh, C.S., Hitte, C., Richman, M., Vignaux, F., Priat, C., Jouquand, S., Werner, P., André, C., DeRose, S., and Patterson, D.F. et al. 2000. An integrated linkage-radiation hybrid map of the canine genome. *Mammalian Genome* **11**: 120-130.

Murphy, W.J., Sun, S., Chen, Z., Yuhki, N., Hirschmann, D., Menotti-Raymond, M. and O'Brien, S.J. 2000. A radiation hybrid map of the cat genome: implications for comparative mapping. *Genome Res* **10**: 691-702.

O'Brien, S.J., Eisenberg, J.F., Miyamoto, M., Hedges, S.B., Kumar, S., Wilson, D.E., Menotti-Raymond, M., Murphy, W.J., Nash, W.G., and Lyons, L.A. et al. 1999. Genome maps 10. Comparative genomics. Mammalian radiations. Wall chart. *Science* **286**: 463-478.

O'Brien, S.J. and Stanyon, R. 1999. Phylogenomics. Ancestral primate viewed. *Nature* **402**: 365-366.

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N. Saito, R., and Suzuki, H. et al. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**: 563-573.

Parker, H.G., Yuhua, X., Mellersh, C.S., Khan, S., Shibuya, H., Johnson, G.S., and Ostrander, E.A. 2001. Meiotic linkage mapping of 52 genes onto the canine map does not identify significant levels of microrearrangement. *Mamm Genome* **12**: 713-718.

Priat, C., Hitte, C., Vignaux, F., Renier, C., Jiang, Z., Jouquand, S., Cheron, A., Andre. C., and Galibert, F. 1998. A whole-genome radiation hybrid map of the dog genome. *Genomics* **54**: 361-378.

Sargan, D.R., Yang, F., Squire, M., Milne, B.S., O'Brien, P.C. and Ferguson-Smith, M.A. 2000. Use of flow-sorted canine chromosomes in the assignment of canine linkage, radiation hybrid, and syntenic groups to chromosomes: refinement and verification of the comparative chromosome map for dog and human [In Process Citation]. *Genomics* **69**: 182-195.

Szabo, C.I., Wagner, LA., Francisco, L.V., Roach, J.C., Argonza, R., King, M.C. and Ostrander, E.A. 1996. Human, canine and murine BRCA1 genes: sequence comparison among species. *Hum Mol Genet* **5**: 1289-1298.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., and Holt, R.A. et al. 2001. The sequence of the human genome. *Science* **291**: 1304-1351.

Vignaux, F., Hitte, C., Priat, C., Chuat, J.C., Andre, C., and Galibert, F. 1999. Construction and optimization of a dog whole-genome radiation hybrid panel. *Mamm Genome* **10**: 888-894.

Vinogradov, A.E. 1998. Genome size and GC-percent in vertebrates as determined by flow cytometry: the triangular relationship. *Cytometry* **31**: 100-109.

Waterston, R.H., Lindblad-Toh, K ., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., and An, P. et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520-562.

Wayne, R.K. 1993. Molecular evolution of the dog family. *Trends Genet* **9**: 218-224.

Wayne, R.K., W.G. Nash and O'Brien, S.J. 1987. Chromosomal evolution of the Canidae. II. Divergence from the primitive carnivore karyotype. *Cytogenet Cell Genet* **44**: 134-141.

Yang, F., Graphodatsky, A.S., O'Brien, P.C., Colabella, A., Solanky, N., Squire, M., Sargan, D.R., and Ferguson-Smith, M.A. 2000. Reciprocal chromosome painting illuminates the history of genome evolution of the domestic cat, dog and human. *Chromosome Res* **8**: 393-404.

Yang, F., O'Brien, P.C., Milne, B.S., Graphodatsky, A.S., Solanky, N., Trifonov, V., Rens, W., Sargan, D., and Ferguson-Smith, M.A. 1999. A complete comparative chromosome

map for the dog, red fox, and human and its integration with canine genetic maps.

Genomics **62**: 189-202.

Web Site References

http://www.tigr.org/softlab/assembler/, The Institute for Genomic Research TIGR Assembler 2.0 website

http://genome.ucsc.edu/, University of California, Santa Cruz Genome Bioinformatics website

http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi, Whitehead Institute/MIT Center for Genome Research Primer 3 website

http://www-recomgen.univ-rennes1.fr/doggy.html, UMR-CNRS 6061 Universite de Rennes Canine Radiation Hybrid Project website

http://www.fhcrc.org/science/dog_genome/dog.html, Fred Hutchinson Cancer Research Center Dog Genome Project website

http://www.math.princeton.edu/tsp/concorde.html, TSP/CONCORDE Home webpage

FIGURE LEGEND

Figure 1. Comparative map of HSA1p and CFA 2, 5, 6, 15 and 17

HSA1p and part of HSA1q are symbolized by a vertical bar, graduated every ten megabases. The anchor sites, indicated by lines between HSA1p /HSA1q and markers placed on the canine RH map, allow one to define conserved segments between human and dog (CS I to CS VII). The entire CFAs are symbolized by vertical bars in which colored boxes delineates the human evolutionary conserved segments determined by reciprocal chromosome painting (Breen 1999, Yang 2001). Numbers indicate HSA origin of the conserved segments. The orthologous position of the HSA1p/1q chromosome on RH maps and CFAs (red-colored boxes) is indicated by brackets. Note that canine maps are inverted with respect to their chromosomal positions. For each CFA, the RH map shows the statistical support symbolized by horizontal bars of variable lengths reflecting the five maps automatically delivered by TSP/CONCORDE. In blue at the top of the RH map, a scale of 0 to 100% reflects the confidence level for the position of each marker. In scrambled regions, markers occupying several positions are bracketed in order to narrow the problematic region into smaller intervals. Cumulated distances between RH markers are reported in TSP units at the end of the horizontal bars.

Marker names indicated in red correspond to gene-based markers (Type I); other markers are colored black. Markers in bold indicate genes or non-coding markers that constitute anchor sites. Markers in grey and outside brackets belong to other HSA orthologous regions.

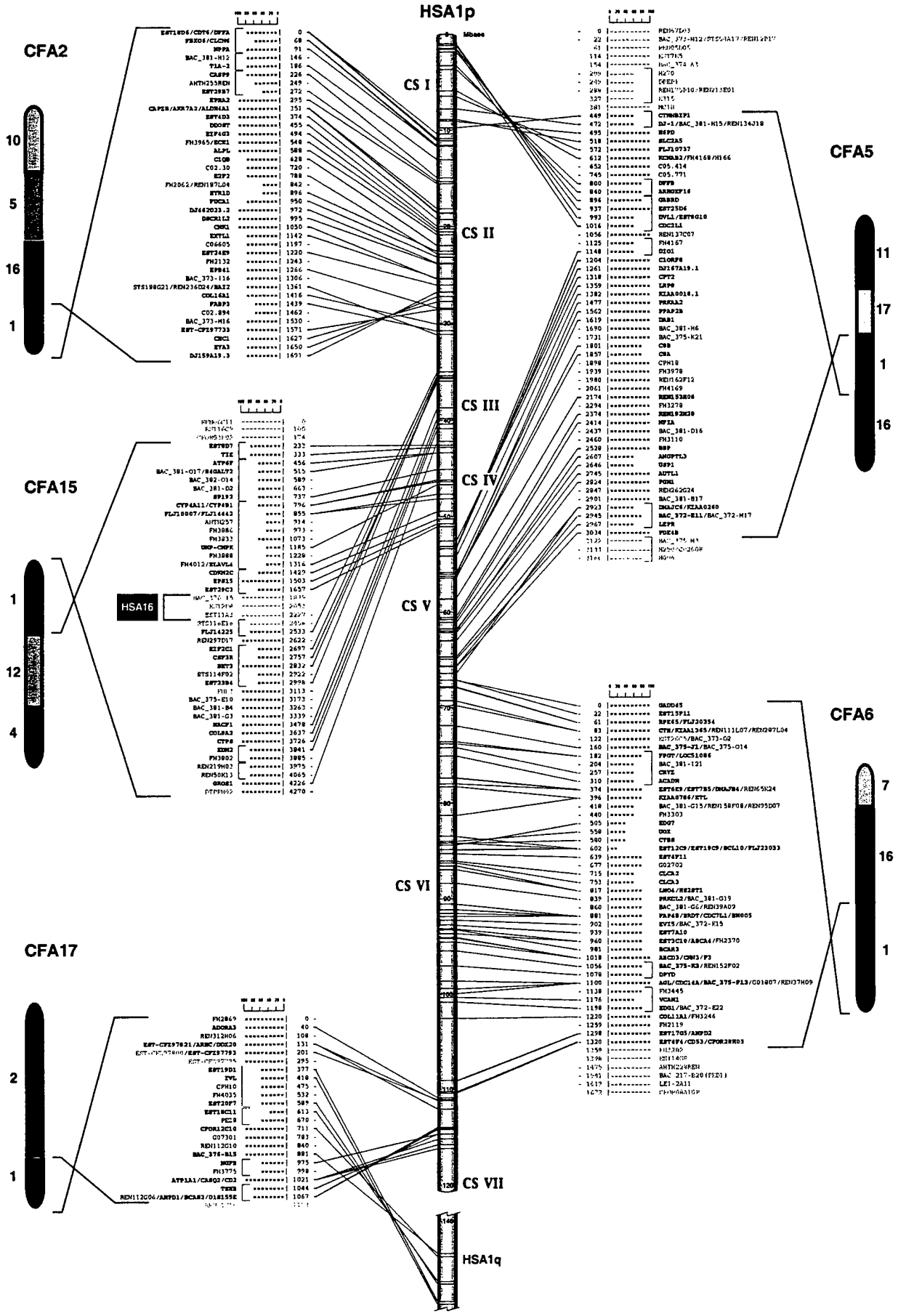
Characteristics of all markers are available at (http://www-recomgen.univ-rennes1.fr/doggy.html) (http://www.fhcrc.org/science/dog_genome/dog.html).

Table 1. Map statistics of conserved segments between HSA1p and canine chromosomes

HSA1p	CFA	Canine CS size (TSP units)	Number of canine markers	Number of mapped positions	Number of canine gene-based markers	Human CS limits (Mb)	Human CS size (Mb)	Total number of human genes in the CS [2]	Number of anchor sites	Average distance between anchor sites in Human (Mb)
CS I	5	567	19	14	13	1-9.8	8.8	123	13	0.7
CS V		1886	36	34	21	52.6-66.4	13.8	91	24	0.6
CS II	2	1691	48	39	35	10.4-31.2	20.8	332	35	0.6
CS III	15	1737	20	20	12	35-42.2	7.2	97	10	0.7
CS IV		1424	22	18	14	42.6-51.1	8.7	107	14	0.6
CS VI	6	1320	78	41	52	67.1-110.4	43.3	273	54	0.8
CS VII[1]	17	293	18	10	14	110.9-123.5	12.6	104	13	1.0
TOTAL		8918	241	176	161		115.2	1127	163	0.7

(1) The numbers corresponding to CS VII only refers to HSA1p orthologous markers

(2) Numbers were retrieved from the NCBI web site: http://www.ncbi.nih.nlm.gov/mapview/maps.cgi?org=hum&chr=1


HSA1p
Mbase
CS I
CS II
CS III
CS IV
CS V
CS VI
CS VII
HSA1q
CFA2
CFA5
CFA15
CFA6
CFA17
HSA16

Version: 6/19/03

A Mutation in the Canine *BHD* Gene is Associated with Hereditary Multifocal Renal Cystadenocarcinoma and Nodular Dermatofibrosis in the German Shepherd Dog

Frode Lingaas*[a], Kenine E. Comstock*[b], Ewen F. Kirkness[c], Anita Sørensen[a], Tone Aarskaug[a], Christophe Hitte[f], Michael L. Nickerson[d], Lars Moe[b], Laura S. Schmidt[g], Rachael Thomas[h], Matthew Breen[e], Francis Galibert[f], Berton Zbar[d] and Elaine A. Ostrander[b]

[a]Norwegian School of Veterinary Science, P.O. Box 8146 N-0033 Oslo, Norway

[b] Clinical and Human Biology Divisions, Fred Hutchinson Cancer Research Center, P.O. Box 19024, D4-100, Seattle, WA 98109-1024 USA

[c]The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850 USA

[d] Laboratory of Immunobiology, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702 USA

[e]Dept. of Molecular Biomedical Sciences, College of Veterinary Medicine, North Carolina State University, Raleigh, NC 27606 USA

[f]UMR 6061 CNRS, Génétique et Développement, Faculté de Médecine, 35043 Rennes Cédex, France

[g]Basic Research Program, SAIC-Frederick, Inc., National Cancer Institute Frederick, Frederick, MD, 21702 USA

[h]Oncology Research, Animal Health Trust, Lanwades Park, Kentford, Newmarket, Suffolk, CB8 7UU United Kingdom

*The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Pre Publication Correspondence: Elaine A. Ostrander, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N., D4-100, PO Box 19024 Seattle WA, 98109-1024. Email: eostrand@fhcrc.org; Phone: 206-667-6979; FAX: 206-667-6396

ABSTRACT

Hereditary multifocal renal cystadenocarcinoma and nodular dermatofibrosis (RCND) is a naturally occurring canine kidney cancer syndrome that was originally described in German Shepherd dogs. RCND is characterized by bilateral, multifocal tumors in kidneys, uterine leiomyomas and nodules in the skin consisting of dense collagen fibers. We previously mapped RCND to canine chromosome 5 (CFA5) with a highly significant Lod score of 16.7 (theta = 0.016). We have since narrowed the RCND interval following selection and RH mapping of canine genes from the 1.3x canine genome sequence. These sequences also allowed for the isolation of gene-associated BACs and the characterization of new microsatellite markers. Ordering of newly defined markers and genes with regard to recombinants localizes RCND to a small chromosomal region that overlaps the human Birt-Hogg-Dubé locus, suggesting the same gene may be responsible for both the dog and the phenotypically-similar human disease. We herein describe a disease-associated mutation in exon 7 of canine *BHD* that leads to the mutation of a highly conserved amino acid of the encoded protein. The absence of recombinants between the disease locus and the mutation in U.S. and Norwegian dogs separated by several generations is consistent with this mutation being the disease-causing mutation.

INTRODUCTION

Canine Hereditary Multifocal Renal Cystadenocarcinoma and Nodular Dermatofibrosis (RCND) is a naturally occurring inherited cancer syndrome in German Shepherd dogs that was first described in 1985 [Lium, 1985 #552; Moe, 1997 #581]. The syndrome is characterized by bilateral, multifocal tumors in kidneys and numerous firm nodules, consisting of dense collagen fibers in the skin and subcutis. In addition, all females examined at an appropriate age have shown uterine leiomyomas and approximately 50% of dogs experience metastasis [Moe, 1997 #581]. Analysis of canine families with RCND strongly indicates an autosomal dominant pattern of inheritance [Lium, 1985 #552; Moe, 1997 #581]. Using a large resource family of Norwegian dogs, we previously mapped RCND to canine chromosome 5 (CFA5) with a highly significant Lod score of 16.7 (theta = 0.016) [Jonasdottir, 2000 #1851].

RCND has some similarities to several human cancer syndromes. A number of provocative genes based upon their phenotype were investigated as possible candidates including the *TSC1, TSC2, TP53, PDK1, KRT9, WT1, FH* and *NF1* genes. Nevertheless, all of these genes have been eliminated based upon their location in the canine map [Breen, 2001 #2247; Guyon, 2003 #2611; Werner, 1999 #546; Priat, 1998 #531; Jonasdottir, 2000 #1608; Jonasdottir, 2000 #1851].

We describe herein the mapping of the RCND locus to a region of CFA5 corresponding predominantly to human chromosome (HSA) 17p11.2. During the course of this work, a human renal cancer syndrome called Birt Hogg Dubé (BHD) that shows some similarity to RCND was mapped to 17p11.2 and the disease-associated gene, termed *BHD*, was subsequently cloned [Birt, 1977 #2642; Khoo, 2001 #2235; Schmidt, 2001 #2236, Nickerson, 2002 #2443]. In addition, a rat model for hereditary renal cell carcinoma was described and the gene responsible was mapped to a portion of rat chromosome 10 that also corresponds to HSA 17p11.2 [Hino, 1993

#570; Hino, 1994 #569]. The function of the protein folliculin, encoded by the *BHD* gene, is unknown. Because of the similarity in phenotype and the corresponding locations in the human and canine genomes, we cloned and then searched for disease-associated mutations in the canine orthologue of the *BHD* gene.

RESULTS

Construction of a high density RH and linkage map of CFA5

Previous work by us localized the RCND locus to CFA5 based upon a maximum LOD score of 16.7 (theta of 0.016) at the marker CO2608 [Jonasdottir, 2000 #1851]. Using whole-chromosome paint probes, evolutionarily conserved chromosome segments between the canine and the human genomes were identified suggesting that CFA5 contains several conserved segments corresponding to portions of HSA 11q, 17p, 1p and 16q [Thomas, 1999 #2655; Breen, 1999 #2217; Yang, 1999 #814; Sargan, 2000 #1991]. By low density radiation hybrid (RH) mapping, C02608 originally appeared to lie in the region close to the boundary between HSA 17p and 1p [Jonasdottir, 2000 #1851].

A high density RH map including 41 microsatellite markers, 10 BACs and 59 genes and an integrated linkage map including 18 markers were constructed as a first step to narrowing the critical region (Figure 1). To accomplish this, the human genome sequence assembly was scanned using the University of California Santa Cruz Human Genome Project Working Draft (http://genome.ucsc.edu/) for genes located on HSA 1p and HSA 17p. These sequences were used to scan the canine 1.3x sequence for orthologous sequences. The resulting sequence reads represented partial sequence of genes from within the region of interest and were used in two ways. First, primers were constructed to directly map the available portion of each gene onto the

canine RH map (Table 1). The ability to order the genes in relation to the recombinants allowed the reduction of the number of candidate genes in the region of interest. Second, the same canine gene sequences were used as probes to screen an 8x canine BAC library for large genomic clones from the region of interest [Li, 1999 #705]. Such clones would be expected to contain additional genomic sequence surrounding each gene and could be used to isolate both additional gene sequence as well as potentially useful microsatellite markers. BACs were pooled and used to construct mini-libraries that were then screened for microsatellites in the region of interest. These microsatellites were placed on the RH map and the markers that were polymorphic in the founder dog were used to fine map recombinants as described below (Table 2).

The minimum RCND recombinant interval includes the canine *BHD* gene

Twenty-six markers in the founder dog were found to be polymorphic and were thus used to analyze the pedigree for additional recombinants in the region of linkage (Figure 2). Further genotyping and haplotype analysis of the Norwegian RCND family identified a recombination in the proximal marker FH4160 that eliminated all genes centromeric to this marker as candidates. We also identified a recombination in the distal marker FH4442 by genotyping and subsequent haplotype analysis that eliminated all genes telomeric to this marker as candidates. The corresponding interval on the human map was inclusive of the genes *GLP2R* (NM_004246, 10.8 Mb) and *MAP2K3* (NM_002756, 22.8 Mb) and spans 12 Mb on HSA 17p. This interval contains the human *BHD* gene (NM_144606, 18.5 Mb) and 85 other genes (RefSeq genes, UCSC Genome Browser on the Human April 2003 Freeze).

Sequence analysis of the canine *BHD* gene identified a disease-associated mutation

Portions of the orthologous canine *BHD* gene were obtained by screening the canine 1.3x sequence database with the sequence of the human gene. This strategy yielded 44 sequences that either encompassed or were within 1 kb of all *BHD* exons. The intron/exon structure of the canine *BHD* gene was deduced from the structure of human *BHD* [Nickerson, 2002 #2443]. Intronic primers were designed to amplify all exons, except exon 1, using DNA from a healthy male Standard Poodle. For exon 1, an untranslated exon, cDNA was isolated from an unaffected Beagle kidney to obtain the sequence near the 5' end of the mRNA. All the canine sequence obtained was compared to the human *BHD* sequence.

All exons were sequenced in three affected dogs and three unaffected dogs from the Norwegian family. In all affected dogs from the family and none of the unaffected dogs, an adenine to guanine mutation in exon 7 was detected (Figure 3). This nucleotide change confers a histidine to arginine mutation in the expressed protein. No missense, nonsense or deletion mutations were found in any other exon segregating with affected dogs.

We next tested 12 RCND-affected German Shepherds from Norway and three from the United States, none of which were descendants of the founder of the Norwegian pedigree. Significantly, the mutation in exon 7 was detected in all 15 RCND-affected dogs. The exon 7 mutation was not detected in 264 unaffected dogs including 63 unrelated, unaffected German Shepherds, 28 Labrador Retrievers, 13 English Setters, 18 Golden Retrievers, 23 Norwegian Elkhounds, 10 Flat-coated Retrievers, 15 Pitbull Terriers, 20 Rottweilers, 16 Boxers, eight Newfoundlands, three Bernese Mountain Dogs and a single dog from each of 47 other breeds. Exon 7 was also examined in a single wolf revealing a sequence identical to the unaffected dogs.

In addition, expression patterns of canine *BHD* in five affected and eight unaffected dogs from the Norwegian pedigree were compared by Northern blot experiments. We saw equivalent

levels of expression of an approximately 3.8 kb transcript and no smaller transcripts in kidneys of both affected and unaffected dogs when Northern blots were probed with *BHD* exon 6-9 (Figure 4).

Expression patterns of canine *BHD* were investigated by Northern blot experiments. We saw expression of an approximately 3.8 kb transcript and no smaller transcripts in unaffected adult canine lung, muscle, skin, kidney, heart, colon, brain and uterus when Northern blots were probed with *BHD* exon 5 (data not shown).

Conservation of the *BHD* amino acid sequence

The folliculin protein is highly conserved across species. Full-length homologues of the human protein (NP_659434) are encoded by the genomes of mouse (NP_666130), rat (XP_220518), *Drosophila melanogaster* (NP_648090), *Caenorhabditis elegans* (NP_495422), and *Schizosaccharomyces pombe* (NP_595962). In addition, gene fragments that are homologous to exons 7 and 8 of the human folliculin gene have been obtained from another mammal (*Bos taurus*; BE481158), a bird (*Gallus gallus*, BG712454), two fish (*Danio rerio*, AL923165; *Oryzias latipes*, BJ487768), a sea squirt (*Molgula tectiformis*, AU281864) and another insect (*Anopheles gambia*, EAA04758). For each of these species, the predicted protein-coding sequence in the region of the canine mutation was aligned (Figure 5). Without exception, all genes and gene fragments encode a His residue at the location of the canine mutation.

A shared haplotype is present in affected Norwegian dogs and distantly related American dogs

Haplotypes were determined in a subset of the Norwegian dogs from the family and the two dogs from the United States with available pedigrees that were diagnosed with RCND. All of the RCND-affected dogs tested shared the haplotype and have the exon 7 mutation. These dogs were genotyped with markers surrounding the RCND locus (FH4229, FH4406, FH4442, FH4464). All the affected dogs share a four marker haplotype spanning approximately 25 cM (Figure 6).

The number of generations between the Norwegian proband and two of the American dogs for whom pedigrees were available, through a common affected ancestor, can be predicted. However, some uncertainty remains due to the high number of common ancestors, some missing pedigree information, and the lack of disease records in the population of German Shepherds. The shortest possible distance between the Norwegian proband and one of the American dogs is 8 generations. However, following the pedigrees through the most likely common ancestors due to accumulation of a number of other affected Norwegian dogs in these lines, they are separated by approximately 12-14 generations. The distance between the Norwegian proband and the other American dog is approximately 22 generations. The two American dogs are separated by at least 10 generations (Figure 6).

In addition, most of 85 other dogs diagnosed by us as having RCND by pathology [Moe, 1997 #581] and where good pedigrees were available, could be traced back to the same pedigrees. Unfortunately, DNA samples are not available to test these dogs for the H255R mutation described here.

DISCUSSION

We have identified a canine gene, *BHD*, which may play a critical role in the pathology of an inherited cancer syndrome in German Shepherd dogs. Identification of gene mutations in human families with BHD disease, together with the high level of identity observed between the *BHD* homologues in divergent species, implies a critical functional role for the folliculin protein. In the German Shepherd Dog, we observed a disease-associated mutation (H255R) in the canine *BHD* gene that confers an amino acid change in a highly conserved region of the protein. It is often difficult to determine if a given missense change is actually disease-causing rather than simply disease-associated in the absence of detailed functional information about the protein. Indeed, while many disease-associated mutations have been reported for cancer susceptibility genes, such as *ATM*, *BRCA1*, and *BRCA2* [Deffenbaugh, 2002 #2652; Boultwood, 2001 #2651], only a subset are confirmed as being disease-causing [Hayes, 2000 #2491; Lavin, 1997 #2653].

In the case of canine *BHD*, three lines of reasoning suggest that the H255R mutation is responsible for RCND. First, evolutionary analysis demonstrates a high level of amino acid sequence conservation between multiple species across exon 7, which contains the H255R mutation. This indicates that this region of the protein is likely to be of functional significance. Specifically, we observed no amino acid differences in H255 in any of 12 species ranging from *H. sapiens* to *S. pombe*. The future availability of functional assays, such as a binding assay to show interactions with other proteins, would allow us to definitively test the biological implications of the H255R mutation.

Secondly, while affected Norwegian and U.S. dogs are separated by at least eight generations, we did not observe the H255R mutation in 264 unaffected dogs of 58 breeds originating from both Norway and the US. Significantly, the H255R mutation was not observed in 63 unaffected German Shepherd dogs, the majority of which were from Norway.

Thirdly and perhaps most compellingly, we found the same H255R mutation in RCND-affected dogs in both the U.S. and Norway and showed that all affected dogs share a common haplotype of four markers spanning approximately 25 cM based upon the canine RH map (Figure 1). The presence of a shared haplotype among all RCND-affected dogs from U.S. and Norway, known to be separated by several generations, makes a strong argument for a founder event. Founder effects are common in dog breeds, resulting when popular sires carrying undetected disease alleles are repeatedly bred into multiple lines within the breed [Ostrander, 2000 #2435]. At the very least, the presence of a shared haplotype among affected individuals argues that if the H255R mutation is not responsible for the disease, another mutation in the shared haplotype that is in linkage disequilibrium with the H255R mutation is. Given that consideration, we can not formally rule out the possibility that there are additional disease-associated mutations in a very closely linked gene or a *BHD* intron or regulatory region. However Northern blot analyses using total RNA from affected and unaffected dogs revealed no apparent differences in expression levels, which argues that message levels and stability are unaffected in RCND dogs. This eliminates the second, but not the first possibility.

We focused our search for candidate genes utilizing both map position and predicted phenotype data. Our ability to map the gene associated with RCND to a small interval was of great importance when selecting *BHD* as the most likely gene. While the region of minimal recombination contains 85 predicted genes based upon comparison with the human sequence, the fact that RCND shares common features with several related syndromes allowed us to limit our search for candidate genes significantly. For instance, human tuberous sclerosis complex (TSC), is similar to RCND except that the latter includes skin tumors and lacks vascular neoplasms. [Roach, 1998 #698; Franz, 1998 #699]. Mutations in the gene encoding fumarate hydratase (*FH*)

cause a predisposion to uterine leiomyomas, benign tumors of the skin and papillary renal cell carcinoma, a phenotype that is strikingly similar to RCND [Tomlinson, 2002 #2648; Toro, 2003 #2654]. However, both *TSC1* and *FH* were definitively eliminated as candidate genes by their canine map position [Jonasdottir, 2000 #1851].

While the phenotypes of RCND and BHD syndrome are quite similar, they are not precisely identical. Human BHD syndrome shows similarity to RCND in that affected individuals experience firm nodules in the skin and subcutis and kidney tumors. Unlike RCND-affected dogs, however, BHD-affected humans frequently experience pneumothorax and do not experience uterine leiomyomas. In addition, there are distinct differences in the types of skin and kidney tumors that occur in the two hereditary syndromes. In BHD, the skin tumors are hamartomas of the hair follicle termed fibrofolliculomas, composed of elongated, delicate epithelial strands in a dense stroma. RCND-affected dogs do not present with hamaratomas, do not show the strands of epithelial cells, and the hair follicles are generally not involved. It is difficult to compare histologic types in affected humans and dogs, nevertheless, in both species the tumors are adenocarcinomas originating from epithelial tubular cells [Lium, 1985 #552].

One striking feature of RCND is that renal tumors were observed in %100 of autopsied affected dogs in the RCND Norwegian pedigree, as well as in seven affected dogs followed clinically over an age of 10-11 years [Moe, 1997 #2644; Moe, 2000 #2645]. By comparison, renal tumors are reported in about 15% of BHD-affected humans [ZBar, 2002 #2647], although differences between the occurrence of renal tumors in BHD-affected humans and RCND-affected dogs could be due to differences between diagnostic methods.

Finally, we found none of the mutations in RCND-affected dogs that have been observed in the BHD-affected human families described previously [Nickerson, 2002 #2443; Khoo, 2001

#2235]. Interestingly, the hypermutable C_8 tract in human *BHD* is interrupted in the center by an "AT" dinucleotide pair in dogs, possibly explaining why the insertion/deletion mutations in this tract that comprised 44% of the *BHD* mutations observed in humans were not seen in any of the RCND-affected dogs. Likewise, in the mouse, the hypermutable C_8 tract is interrupted in the center by a "TG". As additional data becomes available from human families, it will be of interest to see if the differences observed in phenotype between BHD-affected humans and RCND-affected dogs can be correlated with specific genotypes.

Our work described here constitutes the first example of human and canine inherited cancer syndromes with similar phenotypes displaying disease-associated mutations in a both a human gene and its canine orthologue. This particular example focused on kidney cancer, but we hypothesize that similarly structured studies could be used to map other cancer susceptibility genes as well. Many of the same cancers that occur in humans are observed at a very high frequency in certain dog breeds [Ostrander, 2000 #830; Arnesen, 2001 #2643]. Breed associated cancers are observed for Boxers, and Pointers [Dorn, 1987 #805], (lymphoma), Airedale Terriers and Golden Retrievers (soft tissue tumors) [Priester, 1971 #672], Scottish Terriers (melanoma) [Theilen, 1987 #671], Scottish Deerhounds and Rotweillers (osteosarcoma) and Sky Terriers (breast cancer). By utilizing the advantages of canine families and homogenous breed structure, together with the now well developed canine genome map [Guyon, 2003 #2611], we hypothesize that genes involved in both human and canine cancer biology can be mapped. This sets the stage for future studies involving canine pedigrees aimed at mapping and cloning genes for complex human diseases that have not been tractable through the study of large numbers of small human families.

MATERIALS AND METHODS

Canine pedigree development, phenotypic assessment and sample collection

A Norwegian canine colony segregating RCND was established by breeding a single affected male German Shepherd/Flat-coated Retriever to one unaffected female German Shepherd and five unaffected female English Setters [Jonasdottir, 2000 #1851]. All females were unrelated to the male. The five female English setters were related to each other. Offspring were examined for the presence of multiple microscopic renal cysts by exploratory laparotomy, kidney biopsy or necropsy and subsequent histologic examination as described [Moe, 2000 #563]. Blood samples were also drawn from all dogs in the pedigree.

Samples from German Shepherd dogs affected with RCND residing in the United States were obtained by sending requests to veterinarians throughout the country. Once dogs were identified, their owners were contacted asking for participation in the study. Blood samples were drawn from the dogs by their own veterinarians and sent for DNA isolation. Blood samples from control dogs of all breeds were donated by their owners and collected by their own veterinarians either in the United States or Norway.

The Norwegian Animal Research Authority approved colony development, maintenance and sample collection in Norway. Canine blood and DNA samples in the United States were handled as specified by the Fred Hutchinson Cancer Research Center Institutional Animal Care and Use Committee.

Genomic DNA isolation

Genomic DNA was isolated from EDTA-anticoagulated whole blood using standard procedures [Bell, 1981 #448]. All DNA samples were resuspended in 10 mM Tris-Cl (pH 8.0), then quantitated by spectrophotometry.

Partial canine gene sequences were obtained

The 1.3x canine genome sequence used for this study was originally obtained from plasmid libraries of small- (2 kb) and medium-sized (10 kb) inserts, prepared and sequenced at Celera Genomics as described previously for the human genome [Venter, 2001 #2142]. The finished sequence data consists of 6.2 million reads (average read length, 576 bases), representing approximately 1.3x coverage of the haploid canine genome (2.8 Gb) [Vinogradov, 1998 #2237].

The human genome sequence assembly was scanned to identify genes located on HSA 1p and 17p using the University of California Santa Cruz Human Genome Project Working Draft (http://genome.ucsc.edu/). To obtain the corresponding partial canine sequence, the associated human peptide sequence was searched against the complete collection of dog reads using tblastn (W=12). Rarely (~3 % of searches) putative dog orthologues were detected by using less stringent parameters. For each peptide, all homologous dog reads that were identified by the blast searches were assembled at high stringency (99% nucleotide identity) using TIGR Assembler (http://www.tigr.org/softlab/assembler/). Each assembly, or unassembled read, was then searched back against the Ensembl (release 1.1) collection of confirmed cDNAs and peptides (using blastn and blastx, respectively). If the assembly was most similar (at both the DNA and protein levels) to the gene that was used originally for searching, the assembly was considered a fragment of a putative orthologue.

Construction of DNA minilibraries and screening for microsatellites

The partial canine gene sequences were also used to probe a canine BAC library with a mean insert size of 155 kb and a 8.1-fold predicted coverage of the canine genome [Li, 1999 #705]. For each gene, BAC filters were probed with PCR products of ≥400 bp, labeled by random primer incorporation and used at a concentration of 10^6 cpm/mL hybridization solution. Up to five filters were hybridized with 10-15 labeled probes simultaneously in a 7.5 cm diameter bottle, washed and then exposed to autoradiography film using standard techniques [Ausebel, 1987 #2255]. The resulting positive clones were picked from the primary BAC library plates.

To construct mini-libraries, the isolated BAC clones were grown in LB with antibiotics using standard protocols and then pooled [Ausebel, 1987 #2255]. BAC DNA was isolated using a Qiagen Large Construct kit and established procedures [Kelley, 1999 #727]. The BAC DNA was partially digested with two 4-bp cutters, *BfaI* and *MseI*, and the resulting fragments were purified and cloned into the unique *NdeI* site in pGEM-5fZ(+/-). The libraries were transformed into DH5-alpha cells, then screened for common canine microsatellites using $(CA)_{15}$, $(GAAA)_{10}$, $(GTAT)_{10}$ and $(CCTT)_{10}$ oligonucleotides as described previously [Francisco, 1996 #1015]. The resulting clones were sequenced using the pUC/M13 forward and reverse primers using an ABI3700 automated sequencer. After BLAST analysis to eliminate clones containing LINE or SINE elements and identify any gene sequence, primers that bracket microsatellite repeats of 12 or greater were selected using the web-based program Primer3 (www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). Primer sequences and product sizes are shown in Table 2.

Radiation hybrid mapping of canine genes, microsatellites and BACs

Radiation hybrid mapping of the canine genes and microsatellites was done on a 3000 rad panel commercially available from Research Genetics, Inc. All reactions were done using a 96-well format with previously published methodologies [Mellersh, 2000 #567; Breen, 2001 #2247]. PCR reactions contained 1X PCR buffer (Bioline Inc., Randolph, MA USA), 0.5 mM dNTPs, 1.5 mM $MgCl_2$, 0.01 U Taq polymerase (Biolase, Bioline Inc., Randolph, MA USA), 0.3 μM forward and reverse primers and 50 ng of DNA from the panel in a final volume of 15 μl. Reaction conditions were typically as follows: 7.5 min at 95°C, 20 cycles of 94°C for 20 s, 61°C less 0.5°C each cycle to 51°C, 74°C for 20 sec, 10 cycles of 94°C for 20 sec, 51° C for 20 sec, 74°C for 20 s., then one cycle of 74°C for 2 min. Reactions were then held at 4°C. Products were resolved on 1.8% agarose gels electrophoresed for 30 min. Results were visualized under UV light after ethidium bromide staining, photographed, then scored as present, absent or ambiguous. All markers were run on gels at least in duplicate.

Markers were assigned to linkage groups with a Lod cut off score of 8.0 with the program RadMap from the MultiMap computer package [Matise, 1994 #1250; Breen, 2001 #2247]. RH groups containing at least two markers were then ordered using the TSP approach as specified by the CONCORDE computer package (http://www.math.princeton.edu/tsp/concorde.html). TSP/CONCORDE computes five independent RH maps and the resulting maps were subsequently evaluated to produce a consensus map using a method developed by us [Hitte, 2003 #2434]. Inter-marker distances were determined with the rh_tsp_map1.0 version of TSP/CONCORDE that produces map positions in arbitrary TSP units.

Linkage and recombinant mapping

Microsatellites were typed in both members of the RCND pedigree as well as in unrelated affected and unaffected dogs. PCR was performed using 5' - Cy5 labelled primers as reported previously [Jonasdottir, 2000 #1608]. The PCR products was analysed on an ALFexpress ® sequencher (Amersham) with software for Fragment analysis (Allelinks ®).

RCND is assumed to be inherited in an autosomal dominant manner and is fully penetrant. Using the PREPARE option of the Multimap program, each marker was checked for Mendelian inheritance. Two-point linkage analysis was carried out between RCND and each marker and between each pair of markers using the MultiMap software package and markers were ordered by multipoint analyses. The most likely order and spacing of the markers within the linkage group were calculated using multipoint analysis and the GET-LIKELIHOODS function of Multimap. Maps were constructed with framework markers ordered at odds greater than 1000:1 and all remaining markers ordered at odds greater than 10:1 using MultiMap [Matise, 1994 #1250]

Haplotype sharing analysis

One primer of each primer pair was end-labeled using standard conditions [Maniatis, 1982 #1245]. Amplification was carried out with 5 ng genomic DNA using previously published conditions [Jonasdottir, 2000 #1851]. Primer sequences and product sizes are shown on Table 2. PCR products were separated on 4-6% polyacrylamide gels under denaturing conditions at 55°C, visualized by autoradiography and scored manually.

Cloning and sequencing of the canine *BHD* gene

To obtain the partial canine sequence corresponding to the human *BHD* gene, the associated human gene sequence was searched against the complete 1.3x collection of dog reads. Canine sequence was found within 1 kb of all the corresponding human exons. BLAST (www.ncbi.nlm.nih.gov:80/BLAST/) and Repeat Masker (www.repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker) were used to identify any repeated elements in the sequence. Primers were designed using Primer3 (www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) to flank all exons except exon 1 by at least 40 bp. PCR product sizes ranged from 304 to 1755 bp. Treatment of the PCR products with exonuclease I and shrimp alkaline phosphatase was done prior to sequencing. Sequencing was done using the BigDye kit (Applied Biosystems Inc., Foster City, CA USA) and an ABI3700 or 3730 automated sequencer. The sequence of canine *BHD* exon 7 has been submitted to Genbank (Accession #AY326427). Alignment and comparison of sequences from affected and unaffected dogs was done using the Phred/Phap/Consed software packages [Ewing, 1998 #2304; Ewing, 1998 #2307; Gordon, 1998 #2305].

Mutation detection

The sequence of exon 7 was identified by sequencing of a PCR-product from cDNA using primers from exon 6 to exon 9. After initial identification of a mutation in the end of exon 7 in an affected dog, new primers were designed from the start of exon 7 (Ex7F) to intron 7 (In7R) for the purpose of genomic PCR and mutation detection. The sequencing reaction was performed with a nested primer (Ex7FS) 18 bp downstream of the forward primer. A PCR product of approximately 1500 bp was generated using primers

Ex7F (5'-GAGGCAGAGCAATTTGGTT-3') and In7R (5'–TGTTGGATGATTTTGTGTTTGA-3') and standard protocols for PCR under the following cycling conditions: 95°C for 3 min, followed by 35 cycles each of: 95 °C for 30 s, 60 °C for 45 s and 72°C for 90 s. The sequencing reaction was performed with the ET-terminator kit for MegaBACE (Amersham) in accordance with recommendations from the manufacturer using the sequencing primer Ex7FS 5'-GAGAATGAACACGGCCTTC-3' in a 20 µl reaction mixture containing 8 µl "ET-mix", 2.5 µl PCR product, 1.5 µl sequencing primer (5 pmol/µl) and 8 µl H_20. The sequencing reaction was performed by cycle sequencing with the following protocol: initially 95 °C for 1 min then 29 cycles each of: 95 °C for 20 s, 59 °C for 15 s and 60 °C for 60 s. Sequencing was done using an automated sequencer (Molecular Dynamics MegaBACE 1000).

mRNA isolation, Northern analysis and 5-RACE

Tissues were collected from all dogs shortly post mortem. All tissues were then immediately immersed in liquid nitrogen. RNA was isolated from 50-100 mg canine tissues using TRIZOL reagent (Invitrogen Inc., Carlsbad, CA USA) as recommended by the manufacturer. Total RNA was isolated from tissues of two unaffected adult dogs. In addition, total RNA was isolated from kidneys from three unaffected and three dogs affected with RCND.

A kit was used for the Northern blot procedure (NorthernMax, Ambion Inc., Austin TX, USA). 10 µg of total RNA for each sample was loaded onto 1.0% agarose gels containing formaldehyde and electrophoresed at 5V/cm for approximately 2 hours. Ribosomal RNA was visualized under UV light after ethidium bromide staining to check for possible degradation. RNA was transferred to a nylon membrane (Hybond N+, Amersham, Inc. Piscataway, NJ USA) by capillary transfer, then UV crosslinked.

The blot was prehybridized in ULTRAhyb solution (Ambion Inc., Austin, TX USA) for 30 min at 68°C. A PCR product from exon 6-9 or exon 5 was radioactively labelled and used as probe. The probe was labelled in a reaction mixture containing 0.5 µl PCR product, 1.0 µl cold dNTP, 1.0 µl 20 pmole primer, 0.2 µl Taq polymerase, 2.6 µl 10X PCR buffer, 15.7 µl H_20 (total 22 µl). 4 µl [α^{32}P]dCTP was added and the following protocol was used; initial denaturation at 95°C for 3 min followed by 30 cycles each of 95 °C for 30 s, 58 °C for 20 s, 72 °C for 60 s and 72 °C for 5 min. The radiolabeled DNA probe was added at a concentration of 10^6 cpm per mL to the ULTRAhyb solution. The blot was incubated overnight at either 42°C or 68°C in a roller bottle hybridization oven.

The blot was washed in 2 x SSC at room temperature two times for 5 min each, then washed in 0.1 x SSC at 42°C (exon 5 probe) or 65 °C (exon 6-9 probe) two times for 15 min each. The blot was then exposed to film at -80°C with an intensifying screen overnight for autoradiography.

For 5' amplification of BHD cDNA, the SMART RACE cDNA amplification kit (Clontech, Palo Alto, CA USA) was used. One microgram of total RNA isolated from a Beagle kidney was used. First strand cDNA was made according to the manufacturer's instructions. The cDNA was then specifically amplified according to the manufacturer's instructions using a canine *BHD*-specific primer (5'-CGATGGCATTCATGGTGTCCTGGAG-3'). The resulting PCR product was sequenced as described above.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of US Army grant BAAD AAD 19-01-0658, a Burroughs Wellcome Award for Innovative Genomics, an NCI K05 award (K05 CA 90574), an ACS grant RPG 96-045-04-MGO and the American Kennel Club Canine Health Foundation to EAO. The work was supported by grants from the Norwegian Research Council (110642/120 and 153544/140) to F.L. This publication has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under Contract No.N01-C0-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U. S. Government.

REFERENCES

FIGURE LEGENDS

Figure 1: A comparison of linkage, RH and cytogenetic maps of canine chromosome 5.

The vertical line at the far left represents the linkage map constructed using data from the RCND pedigrees. The distances are given in centiMorgans (cM). Markers placed on both the linkage and RH maps are indicated by dotted lines.

The statistical support for the RH map, shown in the center, is symbolized by horizontal bars of variable lengths reflecting the five maps automatically delivered by TSP/CONCORDE. At the top of the RH map, a scale of 0 to 100% reflects the confidence level for the position of each marker. Distances between RH markers are reported in TSP units between the horizontal bars. Underlined marker names correspond to gene-based markers (Type I); Bacterial artificial chromosome markers are indicated by the "BAC' prefix preceeding the number; all remaining

markers are microsatellites. The box encompassing the region on the RH map from FH4160 to FH4442 indicates the minimal recombinant region.

The entire canine chromosome 5 (CFA5) is symbolized on the right by a vertical bar in which shaded boxes delineate the human evolutionary conserved segments determined by reciprocal chromosome painting [Breen, 1999 #2447]. The orthologous position of DI01 on the RH map andCFA5 is indicated by dotted lines.

Figure 2: The canine pedigrees segregating RCND. Affected dogs are represented with black shading and unaffected dogs are unshaded. Marker names are indicated to the left of each row of genotypes. The genotypes of all markers are shown, but the vertical bar representing the haplotypes in the offspring is only shown for the affected proband's side. For the RCND locus, a "1" indicates the wild-type allele (unaffected) and a "2" indicates the mutant allele (affected).

Figure 3: A. The canine *BHD* nucleotide sequence of exon 7 is shown with the mutation indicated in parentheses. B. Chromatographs showing the nucleotide sequences surrounding the mutation. Arrows 1 and 2 indicate the sequence of a heterozygous affected dog. Arrow 3 indicates the sequence of a homozygous unaffected dog. C. The canine folliculin amino acid sequence showing the H255R mutation.

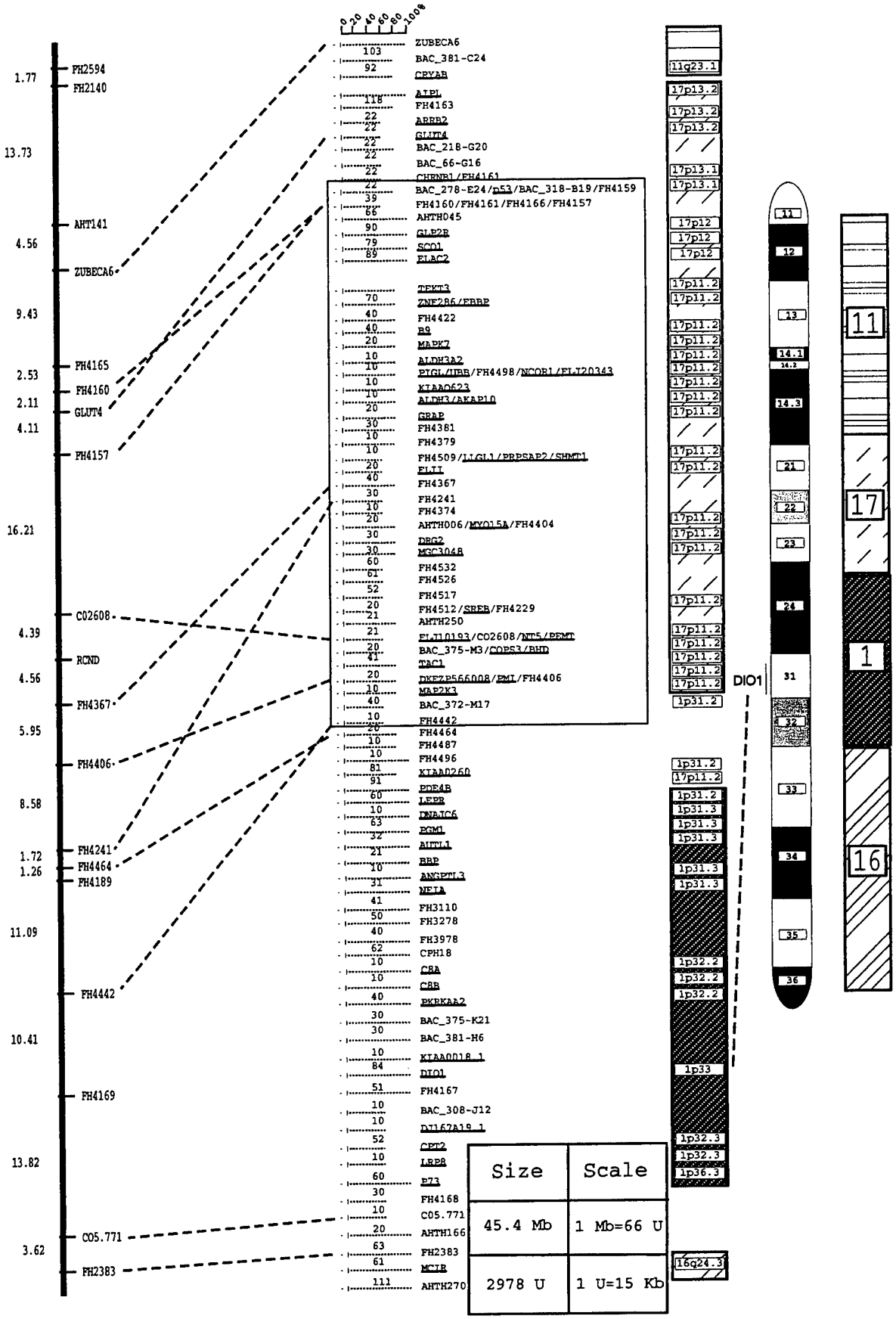
Figure 4: Analysis of *BHD* mRNA levels in RCND-affected and unaffected dogs. A. Northern blot probed with *BHD* exon 6-9. B. Probed with the *GAPDH* gene. The arrows indicate size in Kb.

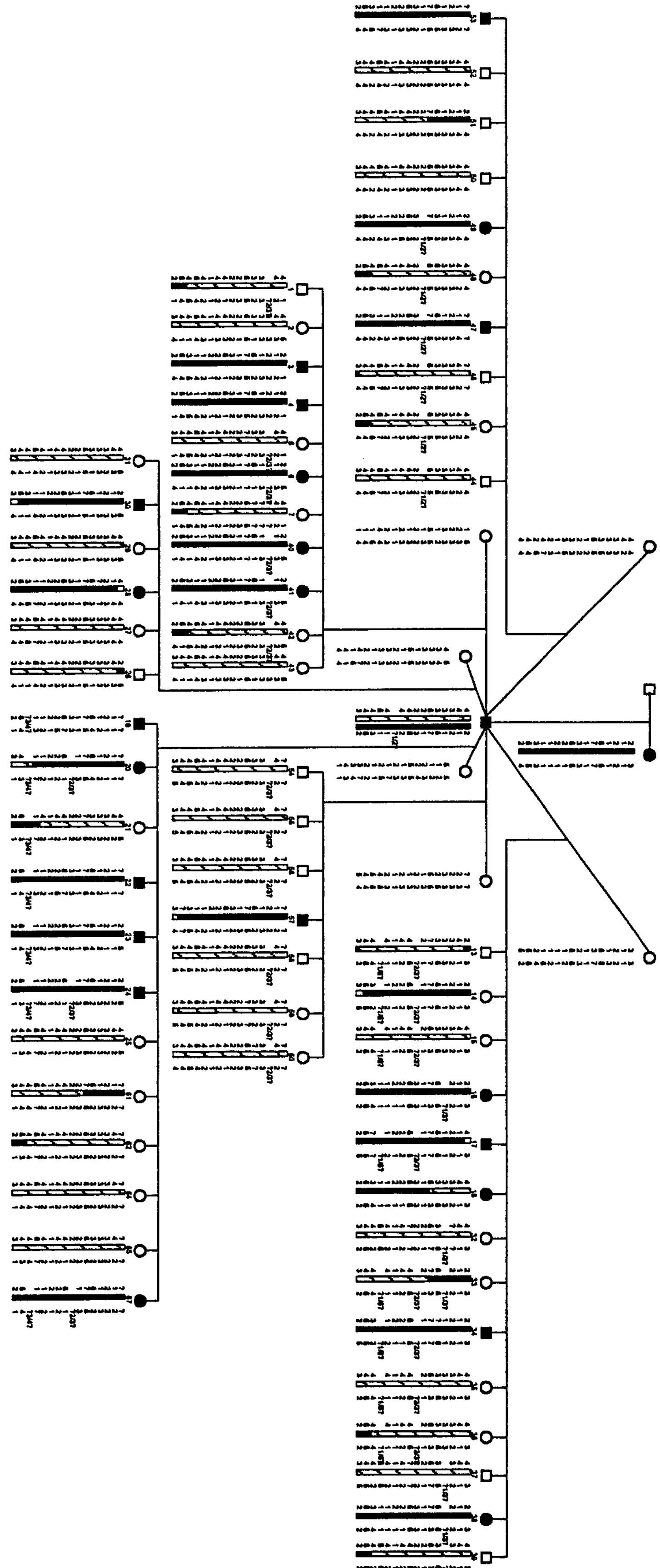
Figure 5: Alignment of folliculin homologues. The arrow indicates the location of the amino acid mutation in RCND-affected dogs. Identical amino acids are in dark gray, conservative differences are in light gray, and nonconservative differences are unshaded.

Figure 6: Haplotype sharing analysis. Affected dogs are black, unaffected dogs are white. The "FH" prefix on the marker names was omitted. Shared haplotypes are boxed and dotted lines indicate more distantly related dogs.

Table 1: Primer sequences and product sizes for RH mapped genes

Table 2: Primer sequences and product sizes for canine microsatellites

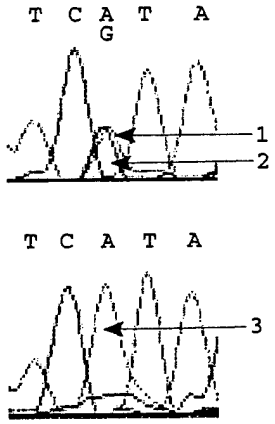

0 20 40 60 80 100%
1.77
13.73
4.56
9.43
2.53
2.11
4.11
16.21
4.39
4.56
5.95
8.58
1.72
1.26
11.09
10.41
13.82
3.62
FH2594
FH2140
AHT141
ZUBECA6
FH4165
FH4160
GLUT4
FH4157
C02608
RCND
FH4367
FH4406
FH4241
FH4464
FH4189
FH4442
FH4169
C05.771
FH2383
ZUBECA6
BAC_381-C24
CPVAB
AIPL
FH4163
ARRB2
GLUT4
BAC_218-G20
BAC_66-G16
CHRNB1/FH4161
BAC_278-E24/p53/BAC_318-B19/FH4159
FH4160/FH4161/FH4166/FH4157
AHTH045
GLP2R
SCO1
ELAC2
TPKT3
ZNF286/ERBP
FH4422
BG
MAPK7
ALDH3A2
PIGL/UBB/FH4498/NCOR1/ELT20343
KIAA0623
ALDH3/AKAP10
GRAP
FH4381
FH4379
FH4509/LLGL1/PRPSAP2/SHMT1
ELII
FH4367
FH4241
FH4374
AHTH006/MYO15A/FH4404
DBG2
MGC3048
FH4532
FH4526
FH4517
FH4512/SREB/FH4229
AHTH250
FLJ10193/CO2608/NT5/PEMT
BAC_375-M3/COPS3/BHD
TAC1
DKFZP566008/PMI/FH4406
MAP2K3
BAC_372-M17
FH4442
FH4464
FH4487
FH4496
KIAA0260
PDE4B
LEPR
DNAJC6
PGM1
AUTL1
BRP
ANGPTL3
NF1A
FH3110
FH3278
FH3978
CPH18
CBA
CBB
PKPKAA2
BAC_375-K21
BAC_381-H6
KIAA0018.1
DIO1
FH4167
BAC_308-J12
DJ167A19.1
CPT2
LRP8
P73
FH4168
C05.771
AHTH166
FH2383
MCIR
AHTH270
11q23.1
17p13.2
17p13.1
17p12
17p11.2
1p31.2
1p31.3
1p32.2
1p32.3
1p33
1p36.3
16q24.3
DIO1
11
17
1
16
Size
Scale
45.4 Mb
1 Mb=66 U
2978 U
1 U=15 Kb



A.

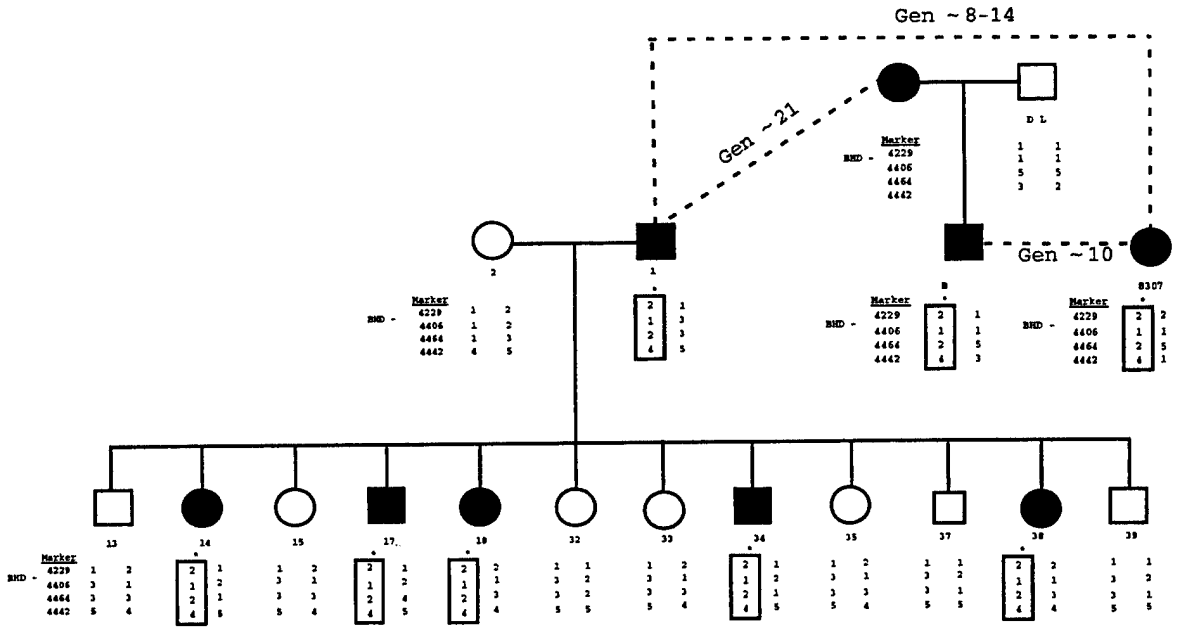
GTATTTGAGGCAGAGCAATTTGGTTGCCCACAGCGTGCCCAGAGAATGAACACGGCTTCACACCAT
TCCTGCACCAACGCAATGGGAACGCAGCTCGTTCACTGACCTCCTTGACAAGCGATGACAACTTGT
GGGCATGCCTTC(A/G)TACCTCCTTTGCTTG

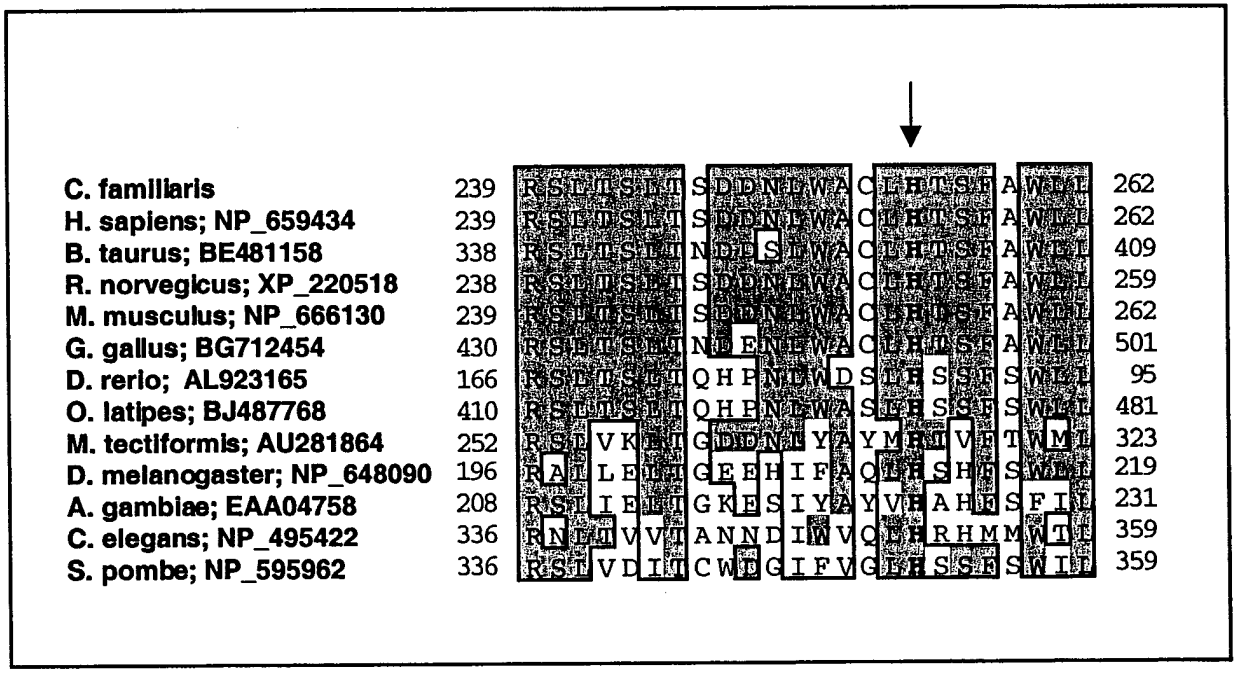
B.

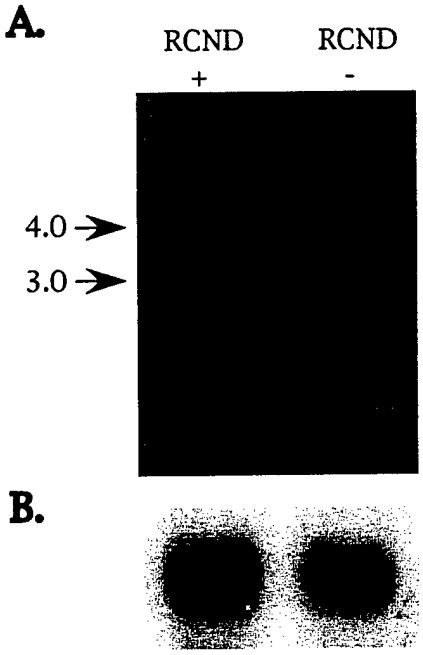


C.

Dog: VFEAEQFGCPQRAQRMNTGFTPFLHQRNGNAARSLTSLTSDDNLWACLHTSFA
RCND dog: VFEAEQFGCPQRAQRMNTGFTPFLHQRNGNAARSLTSLTSDDNLWACL**R**TSFA


Gen ~8-14
Gen ~21
Gen ~10
Marker
4229
4406
4464
4442
BMD -
D L
2
1
8
8307
13
14
15
17
18
32
33
34
35
37
38
39


C. familiaris 239 262
H. sapiens; NP_659434 239 262
B. taurus; BE481158 338 409
R. norvegicus; XP_220518 238 259
M. musculus; NP_666130 239 262
G. gallus; BG712454 430 501
D. rerio; AL923165 166 95
O. latipes; BJ487768 410 481
M. tectiformis; AU281864 252 323
D. melanogaster; NP_648090 196 219
A. gambiae; EAA04758 208 231
C. elegans; NP_495422 336 359
S. pombe; NP_595962 336 359


A.
RCND
+
RCND
-
4.0
3.0
B.

Name	Repeat	Forward Primer (5'-3')
FH2383	Tetra	GACCTGTCTTCTCCTGAGTCTACC
FH3278	Tetra	CTGCTCTTTGTAACCCATGC
FH3978	Tetra	ACCATAGAAGGAATGGTCAGTG
FH4157	Tetra	AATCAAACATAGGCAGTGTGG
FH4160	Di	ACCACAAACACAAATGCTACAG
FH4166	Di	TATGTTTCTTCTTTCCCACCAG
FH4167	Di	GAAGATCATCGTGGGAGATG
FH4168	Di	AGGACCCTTCTCTTATGGAGTC
FH4169	Di	ATTCTGGACAAGTTACTGTGGG
FH4171	Di	AGGAGATGCTACAGGCAGG
FH4229	Di	CTCGTGGAGCTTACCATCC
FH4241	Di	ATGGACCCAGGTTATCTCAGC
FH4367	Di	GCTGGGTATCCACGACTGG
FH4374	Di	AGTGGGAGAGTCTCAGTGTCC
FH4379	Di	GGCTTCAAGCAGATAAAGGAC
FH4381	Di	GCATGAACTTTGTGGAACTGC
FH4404	Di	GGACCGTCAGATTACATGAGC
FH4406	Di	CTCTCATCTATGAAGCATTGTCC
FH4422	Di	TTCTAAAGGGTAGGAATTGAAGG
FH4442	Tetra	GGTTTAGTTTGGTTTTGTTTGG
FH4464	Tetra	CACCTGCCTGGCTTAACA
FH4487	Tetra	AACCACAAGTTTGCTTTTAGC
FH4496	Tetra	GTCTCTGCCTCTGTGTCTCTAT
FH4498	Tetra	GCATGGATGATAAAAGCAACC
FH4509	Di	CCAGTCCACTTGAGTTGCTT
FH4512	Di	TTAGGATATGGAACACCGTGAAC
FH4517	Tetra	GTTCAACACTACAATGATCAAAAGG
FH4526	Tetra	AGTCAGGTGTGAGATCCAGTAGC
FH4532	Tetra	CGCAGGTACACCTTCCTAAACC
CPH18	Tetra	CAGAGATACGTCTTGACACTAGCAGA
ZuBeCa6	Tetra	AGGAGTTACATGCCATAAGCC

Reverse Primer (5'-3')	Annealing temperature (°C)	Product size (bp)
TACCAGAAATTACCTGCCCG	58	500
AATGCCTACCAGGTGAAGG	58	324
TCAGAATCTCTGGGGTCATTAG	58	331
ACGAATCAGCCAGGAGAAGG	58	452
GTTCTCACGCTAGAGAAGGAAG	55	132
CAGGACCTTTATTTCTCATTGG	58	151
TATAGGATGGAGTCTACGGGTG	57/32cyc	340
ACACATGCAGAATGTATCGAAG	55	479
ATTTCCCTGGCCTATAGTTTTC	58	272
CTTTGTGGAATGAAATGTAGGG	58	215
CTGAGGGAGCCCTCTACC	60	374
ATATACGGACTGGGACACTGG	58	203
AGTGGGGAGACCCTGACC	58	357
GTGCTTTCAAGTGTCCTGACC	60	241
GAGCATGGAGCTTGCTTG	60	314
GCTCTCTGTTCCTGAGTGTCC	60	197
ATATACGGACTGGGACACTGG	62	246
ATGGCACTTTTCTGCTTACG	62	168
GGAATAGTCTATGTAATCTCAATGTGC	58	203
CATTCTCAGCCAGGTTTGG	58	359
CTGCCTGATGTTCAGTGTCTT	62	247
ATCTGATTTTCCCATCTCAGG	62	332
CTCCTCAAAGCTTACCCTCA	60	197
TGTGAGTTCTCTATGGCAAAGC	58	284
CCGCCATCTTGAGGAGTT	td 61-51	176
AAGCAGGGTTTGTGTTGTCTG	td 61-51	185
CTAGAGCCTTTCTCAGGTTTGC	60	475
GTGTTTGCTTCATAATCAACAAGG	60	262
AGTTTCTGATTTCAGAGCTCAAGG	60	477
AGCAGACAGTGGGCCATGTT	58	237
CCAGTAAGGATTTTACCAGCC	58	100

Ostrander lab number	Gene	Forward primer (5'-3')
NO647	AIPL OR ALPL	CAGCTACATCAGGGCTTATCC
NO739	AKAP10	TGCTAGCTGGTAGAATTGTGG
NO740	ALDH3	GCTACGTGGACAAAGACTGC
NO684	ALDH3A2	TTGGGTCTATTTCTCTGTGGTC
NO663	ANGPTL3	CTTCAATGAAACTTGGGAAAAC
NO527	ARRB2	GAAGGAGGGWGCCAACAA
NO664	AUTL1	CTTCTCAACAAGCACAAATACG
NO741	B9	GAGAGGGACAGAGGACCTACC
NO665	BBP	TGTGAGGCTACACTGAGTTTTC
	BHD	GCTTTTCTGGGAGGAAAGAGG
NO611	C8A	CATCAATGACTATGGCACTCAC
NO612	C8B	CTCTCAAGCAGACTGACATTTG
NO742	CHRNB1	TGATCTTTGTTCCATTCTACCG
NO685	COPS3	ACAAACCAGCTGACCTCAATAC
NO615	CPT2	TCCAGCTTGAAGTTAAGTCTCC
NO501	CRYAB	AGTTCTCAGCAAGTGGTGCCAGTTCCT
NO617	DAB1	TCCATGAATTATTCTTTCCCAG
NO509	DIO1	ACAGGAGGGCTCCTCAAGTCCT
NO618	DJ167A19.1	GAAAAGGTTGGTGAATAGCTTG
NO686	DKFZ5660084	CAGATAAATGGCCAAAGAAGG
NO666	DNAJC6	GAATTCATTTCCTTTTGCTCTG
NO687	DRG2	CAACATTTGGTTGTATAAGGGG
NO667	E9271	ACGTAACTTTTAACAGCGAAGC
NO688	EBBP	GAGAAGTTGTTTCCAGAAATGC
NO745	ELAC2	AGTTGTTGGGTTTGAATAATGG
NO689	FLII	GAGATCTACTACTGGATTGGCG
NO690	FLJ10193	CATTCTTGGTTCTGACTTGCTC
NO819	FH	TTTCTTTGTCCAAAAGCTAATGC
NO747	GLP2R	ACCTCTACACAGTCCTTTCTGG
NO487	GLUT4	CCGAGATCGAATCCCACTT
NO748	GRAP	GCTGAAGAGATTCTGATGAAGC
NO671	KIAA0018.1	GGGTTGGGAAGTACAAGAAGAC
NO672	KIAA0260	TCAAATAAAATGGATCAGGAGG
NO750	KIAA0623	AGGCTGATCTAAATGGTCTCC
NO502	LEPR	ACGTTTGAGCATCTTTTTATCAAGCA
NO693	LLGL1	AGTGGGGTGCAGTGAAGAG
NO673	LRP8	GAAATTAGCATCTTCCTGTTGC
NO728	MAP2K3	CCTCAAGACACTGGGAGAGG
NO695	MAPK7	TGAGCTCACAATTCTCATTCAC
NO503	MC1R	GACCGCTACCTCTCCATCTTCTACGCG
NO751	MGC3048	CTGATCTGTGGTTAGGGATAGG
NO752	MYO15A	TGAACTTCCTGGTCATTTGG
NO697	NCOR1	AACATTACAGACCGACAACTCC
NO753	NT5M	TCTACAGCCCCTGGTAGAGG
NO674	NFIA	AGTGATGCTGACATTAAGGACC
NO533	P73	CCTTCACCCTCAACTCATCA
NO675	PDE4B	GAAATCTGGAGAGACAGTGAGG

Reverse primer (5'-3')	Product size (bp)	Annealing temperature (°C)
GGAGGCCAGTCTTTAGTAGGAG	491	
CATGGCAGAGGCTAACTGG	324	58
ATTAGATTGCATTCTGCAAGG	413	58
CGAGAAGGACATCCTGGAG	256	58
TATTCATTCATTGTTGCATTCC	164	56
AATTCAATGAGGTTGGTRTCC		
GGTCCTACTAATGCACTCTTGC	255	58
TGCAAGTACTGCTTTGTGTACG	378	60
AATCAATTAAGCTCCCAATTCC	309	56
CAGGACTCAGTCTGGGATGC	700	60
TGTTACCTTATACGGTTTTGCC	212	58
GTGTGTAAACCGCAGACTTCTC	204	58
AAGAAGACTGGCAATTTGTGG	339	60
ATTAAGCAAACACAGGACCAAC	604	60
TTTAACCTCATTCTAGCCAAGG	194	58
TTCCTTGGTCCATTCACAGTGAGGAC	200	58
CAAAGGCATAACAGTTTGTGTG	283	58
CCCCAGCAGTATCCAGTGGG	210	58
AATGGGTGGTAATTCTGAGATG	161	58
AAGTCCTTGAGAGAGGGACTTC	194	60
GGATTCTGGAGTAGCCAATTATAC	126	56
ATCACTGACCTTACCCACAGAG	260	58
TCCCTTCTTTTCTATTCTTCCC	450	56
ATGTTGATTCATACTTGGGAGG	386	58
TATTCCTTCCAAGGTCAGACG	368	58
AAATCTAGTGTGGATGGCTGTC	434	58
AGGCCTAGTAAGGGAACATAGC	275	58
ATCATTTTCCTCTTGTGGTATGG	243	td 64-54
AGGAACTGGAAGTCTTTCTGC	372	58
GGCTGCCCCTTATTTTTAT	187	60
CCCATCACATCAAACTCTGG	396	58
CTAGATAGGCCAAAGATGATGC	494	td 65-55
GTAGGCTTGTTTCACAGGGTAG	253	60
TGGTATTCGCATCAAAATAGG	382	58
AACAGAGGGCTGCCTCCTGCCCTCA	275	58
ATCCGGAAGGAGGACATTAG	227	58
ACCTGGTCCTGACTATCATCTG	307	58
GTGGATTCTGTGGCTAAGACG	350	60
AGCCATGCTTATCTTTCTTGTC	221	58
CAGGCGCGGGCAAGCATGTGGACGTA	220	58
CACGATCAAAGACATCAGTGG	304	58
GGCATGTAGCCAGTAAGAGC	369	58
CAATTACCTTTACTACCACCGC	479	58
CATGATGACTTGCCTTTGG	387	58
GGAACGAGAAATGAAACAAGAG	341	58
AGCCATAGGGATACCTGCTC	430	58
ACCAAGTCGTTGGAATTGTATC	438	58

NO754	PEMT	CATGAGCTTAGGGAGAAATGC
NO676	PGM1	AACTAGCCCTACCTTTTCCAAC
NO755	PIGL	TCTGTGGCCCTTACTCTTCC
NO678	PRKAA2	ATCTCAAAGACATGGTGGGTAG
NO700	PM1	AACTGTCTCCTCTTTCTCCTCC
NO756	PRPSAP2	CTTGCACTGTTTGCTGTAAGG
NO758	SCO1	CTAATCACCCACATCCAAGC
NO759	SHMT1	GCTGGTGAAATTCTCTGACG
NO702	SREB1	TTAATAAAACGCCTTTAGCAGC
NO760	TAC1	CCACAGCATGTTCAAGAAACT
NO703	TEKT3	GAGTGATAAACAGGCAGCCTAC
NO761	UBB	GAAATGGTTTCCCATTACACC
NO682	USP1	TGTTCAACTGAAATGATCAAGG
NO764	ZNF286	TCAGAGCACACATCTTGTTCA

GTTGTCCAGGATGTTGAACG	369	58
GTTCTCCTGTTTTCATGGTCTC	374	58
GTGTCCAGGAATTCACAATCC	357	58
ATCTGGTTCATGAAGGTTTGAC	457	58
GTTTATAGGCGTCGTACTCCAC	408	58
TTCTCTGCCTGATAGTGAAGG	387	58
CAATTGGTGGAGCTTTACTGG	414	58
AACAGGGATGAAAATAACAAGG	396	58
CGCTTCTTCCTGAGTAGTGC	476	58
GTACTGGGATCCCCTGCT	279	58
CTGAGGAATACTCCTGGTTACG	274	58
AACTCTCCACCTGGTTCTCC	412	58
CACAGTCACAGTTGATCGTACC	350	58
GAGCAGCATTTTACGGGTATT	396	58