The DNA sequence and comparative analysis of human chromosome 20

P. Deloukas, L. H. Matthews, J. Ashurst, J. Burton, J. G. R. Gilbert, M. Jones, G. Stavrides, J. P. Almeida, A. K. Babbage, C. L. Bagguley, J. Bailey, K. F. Barlow, K. N. Bates, L. M. Beard, D. M. Beare, O. P. Beasley, C. P. Bird, S. E. Blakey, A. M. Bridgeman, A. J. Brown, D. Buck, W. Burrill, A. P. Butler, C. Carder, N. P. Carter, J. C. Chapman, M. Clamp, G. Clark, L. N. Clark, S. Y. Clark, C. M. Clee, S. Clegg, V. E. Cobley, R. E. Collier, R. Connor, N. R. Corby, A. Coulson, G. J. Coville, R. Deadman, P. Dhami, M. Dunn, A. G. Ellington, J. A. Frankland, A. Fraser, L. French, P. Garner, D. V. Grafham, C. Griffiths, M. N. D. Griffiths, R. Gwilliam, R. E. Hall, S. Hammond, J. L. Harley, P. D. Heath, S. Ho, J. L. Holden, P. J. Howden, E. Huckle, A. R. Hunt, S. E. Hunt, K. Jekosch, C. M. Johnson, D. Johnson, M. P. Kay, A. M. Kimberley, A. King, A. Knights, G. K. Laird, S. Lawlor, M. H. Lehvaslaiho, M. Leversha, C. Lloyd, D. M. Lloyd, J. D. Lovell, V. L. Marsh, S. L. Martin, L. J. McConnachie, K. McLay, A. A. McMurray, S. Milne, D. Mistry, M. J. F. Moore, J. C. Mullikin, T. Nickerson, K. Oliver, A. Parker, R. Patel, T. A. V. Pearce, A. I. Peck, B. J. C. T. Phillimore, S. R. Prathalingam, R. W. Plumb, H. Ramsay, C. M. Rice, M. T. Ross, C. E. Scott, H. K. Sehra, R. Shownkeen, S. Sims, C. D. Skuce, M. L. Smith, C. Soderlund, C. A. Steward, J. E. Sulston, M. Swann, N. Sycamore, R. Taylor, L. Tee, D. W. Thomas, A. Thorpe, A. Tracey, A. C. Tromans, M. Vaudin, M. Wall, J. M. Wallis, S. L. Whitehead, P. Whittaker, D. L. Williams, S. A. Williams, L. Wilming, P. W. Wray, T. Hubbard, R. M. Durbin, D. R. Bentley, S. Beck & J. Rogers

The Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK

The finished sequence of human chromosome 20 comprises 59,187,298 base pairs (bp) and represents 99.4% of the euchromatic DNA. A single contig of 26 megabases (Mb) spans the entire short arm, and five contigs separated by gaps totalling 320 kb span the long arm of this metacentric chromosome. An additional 234,339 bp of sequence has been determined within the pericentromeric region of the long arm. We annotated 727 genes and 168 pseudogenes in the sequence. About 64% of these genes have a 5' and a 3' untranslated region and a complete open reading frame. Comparative analysis of the sequence of chromosome 20 to whole-genome shotgun-sequence data of two other vertebrates, the mouse *Mus musculus* and the puffer fish *Tetraodon nigroviridis*, provides an independent measure of the efficiency of gene annotation, and indicates that this analysis may account for more than 95% of all coding exons and almost all genes.

The finished reference sequence of the human genome is now in sight, underpinned by the recently published working draft^{1,2}. From the outset of the Human Genome Project, the plan has been to determine the complete sequence of each chromosome to an accuracy of greater than 99.99%, and to cover more than 95% of the gene-containing part of the genome (the euchromatin). This finished 'gold' standard was defined and upheld on completion of the first two human chromosomes, 22 and 21 respectively. Here we report completion of the sequence of the first metacentric human chromosome, chromosome 20, to these standards. Analysis of the finished sequence has benefited from comparison with substantial new data sets that were not available at the time of the previous finished chromosome analyses. These include new collections of human and mouse messenger RNA sequences, the protein indices of fully sequenced model organisms, and extensive sequencing of two vertebrates genomes, those of the mouse and the puffer fish T. nigroviridis. As a result, we were able to assess the quality and completeness of human gene annotation by independent analyses. The application of new analytical tools has also enabled assessment of predictive methods to define transcription start sites and other features of gene structures, although these require further development and calibration with the finished annotated sequence.

Clone map and finished sequence

We identified a set of 629 minimally overlapping clones (the tiling path) that spans the euchromatic regions of the short (p) and long (q) arm of human chromosome 20. The tiling path consists of 455 P1-derived artificial chromosomes (PACs), 169 bacterial artificial chromosomes (BACs), 3 yeast artificial chromosomes (YACs), 1 cosmid and 1 polymerase chain reaction (PCR) product (Fig. 1). The euchromatic portion of the chromosome is represented in six

contigs with one contig covering the entire p arm (Table 1). Boundaries between euchromatin and heterochromatin were identified by presence of satellite repeats in the sequence of clones located at the most distal and proximal ends, respectively, of the contigs flanking the centromere, and served as logical termination points for map construction. Clones located at the centromeric boundary of the p arm (Fig. 1) gave an additional signal at 20q11.1 upon fluorescent *in situ* hybridization (FISH) on metaphase chromosomes. We constructed an additional two-clone contig representing this duplication (Fig. 1) and postulate that it is located in the heterochromatic region of the q arm. In contrast to the p arm, four gaps remain in the clone map of the q arm. Three of them are clustered within a 1.2-Mb region at qtel (Fig. 1). We anticipate that the sequences in these gaps are unclonable to the host–vector systems used in this study, probably owing to the high guanine

Table 1 Sequence contigs on chromosome 20							
Size (bp)	Size estimate (kb)						
26,257,626	ND						
5,063,606	20						
24,982,240	~50						
1,147,210	~100						
35,826	~150						
1,700,790	100						
59,187,298 234,339 59,421,637							
	Size (bp) 26,257,626 5,063,606 24,982,240 1,147,210 35,826 1,700,790 59,187,298 234,339						

ND, not determined

and cytosine (G+C) content of the sequence in this region. All four euchromatic gaps were sized by FISH of clones immediately flanking each gap to extended DNA fibres. No gap was estimated to be larger than 150 kb and all the gaps together account for no more than 320 kb of DNA (Table 1). Finally, we defined the location of both telomeres. At the end of the p arm (ptel), clone RP11-530N10 (EMBL accession code AL360078; Fig. 1) ends about 10 kb away from the block of subtelomeric repeats, which extends for 40–50 kb on the basis of the telomeric half-YAC yRM2005 (ref. 3 and H. Riethman, personal communication). A larger allelic variant of the subtelomeric repeat block is also known, half-YAC yA35 (ref. 4). At the end of the q arm (qtel), clone RP11-476I15 (AL137028; Fig. 1) contains part of the subtelomeric repeat block. Each clone of the tiling path was subjected to random subcloning and sequencing. On the basis of internal and external⁵ quality checks, we estimate the accuracy of our finished sequence to exceed 99.99%. Each clone has been finished according to the agreed international finishing standard for the human genome (http://genome.wustl.edu/gsc/ Overview/finrules/hgfinrules.html). In total, we finished 59,421,637 bases in seven sequence contigs. The size of each sequence contig is given in Table 1; the largest one spans the 26,257,626 bp of the entire p arm. The four gaps account for 0.32 Mb (Table 1). Thus, the sequence covers 99.46% of the euchromatic part of chromosome 20, which spans 59.5 Mb. Our estimate for the total size of the chromosome, based on size estimates of 3 Mb for the centromere and 0.2 Mb for subtelomeric repeats, is 62.7 Mb, which is smaller than a previous estimate of 72 Mb (ref. 6).

Gene index of chromosome 20

The finished genomic sequence was first analysed for G+C content and CpG islands. Interspersed and simple tandem repeats in the sequence were then masked and the masked sequence was compared against protein, DNA and expressed sequence tags (ESTs) using BLASTX and BLASTN⁷. In parallel, gene structures were predicted *ab initio* in the masked sequence on a clone-by-clone basis with the programs FGENESH⁸ and GENSCAN⁹.

A total of 895 gene structures was annotated in the finished sequence on the basis of human interpretation of the combined supportive evidence generated during sequence analysis (see Fig. 1). The structures were divided into five groups: (1) 335 'known' genes, that is, those that are identical to known human complementary DNA or protein sequences (all known genes were in the LocusLink database, http://www.ncbi.nlm.nih.gov/LocusLink); (2) 222 'novel genes', that is, those that have an open reading frame (ORF), are identical to human ESTs that splice into two or more exons, and/or have homology to known genes or proteins (all species); (3) 23 'novel transcripts', that is, genes as in 2 but for which a unique ORF cannot be determined; (4) 147 'putative genes', that is, sequences identical to human ESTs that splice into two or more exons but without an ORF; and (5) 168 'pseudogenes', that is, sequences homologous to known genes and proteins but with a disrupted ORF.

Excluding the pseudogenes, chromosome 20 has a gene density of 12.18 per Mb, which is intermediate to 6.71 (low) and 16.31 per Mb (high) reported for chromosome 21 and 22, respectively^{10,11}. We used the gene density of chromosomes 20, 21 and 22 from ref. 12 to adjust the number of genes on each of these chromosomes. The adjusted figures were then used to extrapolate a number of 31,500 genes for the whole genome, which is in agreement with recent estimates^{1,2}.

The analysis of chromosome 20 benefited from the availability of new large data sets to assist the gene annotation. These included human (for example, Genoscope) and mouse ESTs and 'full-length' cDNAs (for example, RIKEN mouse cDNA collection) as well as the protein indices of fully sequenced model organisms such as *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster* and *Arabidopsis thaliana*. Some 81% of the 557 genes in groups 1 and 2 (and 64% of all the annotated genes), have a full ORF as defined by a starting ATG codon and the presence of a 5' and a 3' untranslated region (UTR). Often a stretch of nucleotides immediately preceding the starting ATG seems to be part of the ORF. When the supporting evidence (for example, ESTs) terminated within such a stretch, we did not annotate a 5' UTR. Such genes were also included (8.9% of the 557 genes) in the above set.

The transcription start sites of most genes in the human genome are not yet known. We carried out several analyses to assist the annotation of the 5' ends of as many genes on chromosome 20 as possible. Analysis of the unmasked sequence predicted a total of 660 CpG islands, of which 389 are located near (5 kb upstream or 1 kb downstream) the first exon of an annotated gene structure. Many of the remaining predicted CpG islands have intragenic locations, which in our view does not allow a direct correlation between the observed number of CpG islands and the number of genes on chromosome 20. Among the genes with complete structures, 303 (67%) are associated with a CpG island at their 5' ends, which is in good agreement with the previously reported figure of 60% (ref. 13). We also scanned the sequence of chromosome 20 for putative transcription start (TS) sites using the probabilistic TS site detector program Eponine (T. Down, unpublished). Eponine is optimized for mammalian genomic DNA sequences and detects likely TS sites on the basis of the surrounding sequence (typically 500 bases upstream to 100 bases downstream). Multiple predictions are often clustered, suggesting alternative TS sites for a gene. Eponine has a detection sensitivity of 40%, on the basis of an analysis of human chromosome 22. We found 1,432 TS sites on chromosome 20, of which 492 (34%) are located within 2 kb of the first exon of an annotated gene. In the set of genes with complete structures, 402 TS sites are associated with 166 genes (37.5%) of which 159 (95.8%) have a CpG island at their 5' end. So, Eponine predicts multiple TS sites per gene (the mean value is 2.42 in the 166 genes) and has a bias in predicting TS sites in genes associated with a CpG island at their 5' end.

The 727 genes (that is, introns and exons) extend over a total of 25,213,914 bp (mean 34,682 bp per gene). Excluding expressed pseudogenes, 42.4% of the reported sequence of chromosome 20 is therefore transcribed. Exons account for only 2.43% of the sequence and the mean exon size is 283 bp. A summary per gene group is given in Table 2, which includes figures reported from the

Figure 1 The sequence map of human chromosome 20 and its features. The short (p) and the long (q) arm of the chromosome are depicted in the top and bottom panels, respectively. The features of each chromosome arm are shown from top to bottom as follows: (1) The finished sequence of each clone in the tiling path as a yellow line. Sequence positions are indicated in megabases along the x-axis of 'G+C content' (see 4, below). Eight of the clones were isolated and sequenced elsewhere, namely AC005808 (LBNL H136; BAC 185), AC005914 (LBNL H135; BAC 189), AC006076 (LBNL H133; PAC 12), AC004762 (LBNL H134; PAC 128), AC005220 (LBNL H80; BAC 99), AC004501 (LBNL H144; BAC 121) and AC004505 (LBNL H65; PAC 86C1) at the Joint Genome Institute¹⁶ and AC006198 (RP11-3A1) at the Whitehead Institute (Massachusetts Institute of Technology Center for Genome Research). The centromere has been arbitrarily drawn to span 3 Mb. The exact location of contig AL121762-AL441988 in the heterochromatic region of the q arm is not known. Gaps in the map appear as greenish bars. The width of the bar represents the size estimate obtained by fibre FISH. (2) The location of genetic markers. (3) The distribution of the main types of repeats in the sequence. (4) Plot of the G+C content of the sequence. (5) Plot of the SNP density along the sequence. (6) The location of predicted CpG islands. (7) The location of the annotated gene structures. Right and left coloured arrows indicate gene structures on the + and - strand, respectively. The most 3' end of each gene is drawn halfway along the arrowhead. Only the genes of the annotation group 1 (known; dark blue) and 2 (novel; blue) are named. When no gene symbol is available, the gene name used in the EMBL sequence submission file appears (for example, dJ583P15.4). CDS, protein-coding sequence.

Table 2 Structural charac	teristics of a	annotated gene	structures
Chromosome, gene type	Mean size (kb)	Mean exon size (bp)	Mean number of exons
Chr20, known genes	51.3	294	10.3
Chr21, known genes	57.0		
Chr20, novel genes	25.1	278	5.7
Chr20, putative genes	9.1	217	2.5
Chr21, novel + putative genes	27.0		
Chr20, genes	34.7	283	7.1
Chr21, genes	39.0		
Chr20, pseudogenes	1.9	499	1.4

Structural characteristics of groups of annotated gene structures on chromosome 20 are shown, and compared with similar groups on chromosomes 21 and 22.

292

266

6.0

54

27.6

192

Chr20, all

Chr22, all

analyses of chromosomes 21 and 22 (refs 10, 11). Gene size varies substantially, from 1,234,386 bp (gene C20orf133 (AL117333–AL049633), which is similar to a low-density lipoprotein-related protein, LRP16) to 339 bp (gene C20orf127 (AL121753)). Exon sizes are fairly constant, with the exception of 3' terminal exons (for example, 8,181 bp in *PTPRT*), in contrast to intron sizes, which vary from 33 bp (C20orf97 (AL034548)) to 523,790 bp (*CDH4*).

For 209 (29%) of the annotated genes, we found alternative splice forms. Alternative splicing can, for example, give rise to two distinct peptides by exclusion of the exons encoding a functional domain from one but not the other transcript. The transcript for the soluble form of attractin (ATRN; AL353193-AL132773) lacks the five exons that encode the transmembrane and cytoplasmic domains and are present in the transcript that encodes the membrane form of the protein. Splice variants may encode structurally unrelated peptides. A complex example of alternative splicing and genetic imprinting is found in the GNAS1 locus (AL132655-AL109840). NESP55 and XLAS are transcribed from distinct mono-allelic promoters located upstream of the bi-allelic promoter that drives the transcription of the gene for the α -subunit of the stimulatory guanine-nucleotidebinding protein G_s . $G_{s\alpha}$ is encoded by exons 1–13. A large G protein, XLas, is generated by in-frame splicing of an upstream exon (bp 120,789-121,953; AL132655) to exon 2 (bp 39,075-39,119; AL121917) of G_{sa}. An additional exon located further upstream (bp 106,368–107,508; AL132655) splices again to exon 2 of $G_{s\alpha}$ but not in frame, giving rise to a structurally unrelated peptide, NESP55. An antisense transcript (dJ806M20.3.6; AL132655) postulated to regulate this imprinted region has also been reported¹⁴. In total, we annotated six isoforms of GNAS1. One gene (PLCB4; AL121898-AL031652) was found to have the most isoforms

(eight); in most cases of genes with alternative splicing (130 genes) we observed two isoforms. Typically, we annotated the longest possible terminal exon and did not create entries for alternative splice forms on the basis of alternative polyadenylation sites. If we exclude the putative genes that have mainly incomplete structures, then 35% of the genes (average of 1.65 transcripts per gene) show alternative splicing. This is in agreement with previous estimates¹⁵. Analysis of chromosomes 19 and 22 (ref. 1), both generich chromosomes, showed a higher extent of alternative splicing.

Protein index of chromosome 20

We analysed the proteome of chromosome 20 using InterProScan (http://www.ebi.ac.uk/interpro/scan.html) to look at the distribution of known protein domains. The InterPro database combines information on protein families, domains and functional sites from the databases Pfam, PRINTS, PROSITE, SMART and SWISS-PROT (see http://www.ebi.ac.uk/interpro for links). Of all proteins encoded on chromosome 20, 73.5% have an InterPro match and 30% are multidomain with an average of 2.1 distinct InterPro domains. As shown in Table 3, many of the most frequent domains in the chromosome 20 proteome rank in similar order as in the human proteome¹. There are, however, five domains for which chromosome 20 seems enriched. Four of them-the cysteine proteases inhibitor (IPR000010), the immunoglobulin subtype (IPR003599), the whey acidic protein (WAP)-type 'four-disulphide core' domain (IPR00222), and the pancreatic trypsin inhibitor (Kunitz/Bovine) domain (IPR002223)-are found in proteins encoded by three gene clusters along the chromosome.

Functionally related gene clusters indicate probable ancestral gene-duplication events. The first cluster, at 1.5 Mb (Fig. 1), extends from AL109658 to AL034562 and includes genes with immunoglobulin and immunoglobulin-like domains that are involved in signal transduction and cell adhesion (SIRPB1, SIRPB2 and PTPNS1). The annotated genes PTPN1L and PTPNS1L2 and pseudogenes dJ576H24.1 and dJ673D20.1 are new members of this gene family. Interestingly, the apparently functional gene PTPNS1L2 is located within a larger fragment of 33,048 bp (AL049634 and AL592544), which is an insertion type of polymorphism. The insertion allele is represented in the RP4 PAC library but not the RP11 BAC library. Using a panel of 174 Caucasians, we estimated that the frequency of the insertion allele is 37.3%. The second cluster, at 23.5 Mb (AL096677-AL121831; Fig. 1), comprises members of the cystatin gene family, which encode protease inhibitors with antibacterial and antiviral activities. An additional member, CST7, is located about 1 Mb distal of the main

	InterPro code			Abun	dance			
Rank (genome rank)		Chr20	Hs	Dm	Ce	At	Sc	Name
1 (2)	IPR000822	26	576	341	205	169	53	Zinc finger, C2H2 type
2 (3)	IPR000719	12	481	230	419	1,033	116	Eukaryotic protein kinase
	(IPR002290	12	316	158	219	856	113	Serine/threonine protein kinase family active site)
	(IPR001245	11	193	82	121	477	4	Tyrosine kinase catalytic domain)
3 (ND)	IPR002965	11	190	175	62	178	0	Proline rich extensin
4 (ND)	IPR000010	9	20	4	3	7	0	Cysteine proteases inhibitor
	(IPR003243	8	10	0	1	7	0	Cystatin C and M)
5 (4)	IPR000276	9	373	84	368	0	0	Rhodopsin-like GPCR superfamily
6 (13)	IPR000561	9	234	86	137	42	1	EGF-like domain
7 (24)	IPR000008	8	111	42	53	99	11	C2 domain
8 (ND)	IPR000504	8	203	135	111	248	54	RNA-binding region RNP-1 (RNA recognition motif)
9 (1)	IPR003006	8	584	134	66	0	0	Immunoglobulin and major histocompatibility complex domain
	(IPR003599	7	177	27	11	0	0	Immunoglobulin subtype)
10 (ND)	IPR000345	7	2	4	3	3	3	Cytochrome c family haem-binding site
11 (ND)	IPR001124	7	8	0	10	2	0	Lipid-binding serum glycoprotein
12 (ND)	IPR002221	7	6	4	7	0	0	WAP-type four-disulphide core domain
13 (ND)	IPR002223	6	13	22	38	0	0	Pancreatic trypsin inhibitor (Kunitz/Bovine) family

At the time of analysis, the InterPro database contained 18,149 *Homo sapiens (Hs,* incomplete), 13,843 *Drosophila melanogaster (Dm)*, 18,581 *Caenorhabditis elegans (Ce)*, 25,677 *Arabidopsis thaliana (At)* and 6,176 *Saccharomyces cerevisiae* (Sc) protein entries. Rows in parentheses correspond to InterPro domains that are 'children' (= members) of a broader InterPro domain. ND, not determined. The P-loop motif domain IPR001687 was excluded, owing to low specificity.

cluster (AL035661). Only two of the known cystatins are not on chromosome 20 (IPR003243; Table 3). We annotated three new members (*CST8L*, *CSTL1* and *CST9L*) and two pseudogenes. The third cluster, at 43.5 Mb (AL049767–AL050348; Fig. 1), includes 11 genes that encode proteins with a WAP-type four-disulphide core domain (IPR002221) and/or a pancreatic trypsin inhibitor (Kunitz/ Bovine) domain (IPR002223). A fourth cluster that includes genes for the semenogelins SEMG1 and SEMG2 (semen proteins involved in reproduction) is located within the third gene cluster between members *PI3* and *SLPI*, which have only a WAP-type domain.

Chromosome landscape

The sequence of chromosome 20 has an average G+C content of 44.1%, which is slightly higher than the genome average of 41%. The distribution of the G+C content fluctuates along the chromosome, and regions with higher G+C have a higher gene density (Fig. 1). For example, the sequence from 49.5 to 54 Mb has an average G+C content of 41.1% and a gene density of only 4.9 genes per Mb, in contrast to the region between 60 and 62.5 Mb, which has 56.6% G+C content and a gene density of 28 genes per Mb. Given a sequence length and gene ratio of 1.25 and 1.65, respectively, between the q and the p arm, the q arm seems rich in genes. Gene density can drop as low as 1.54, for instance in a 1.9-Mb region between AL136990 and AL139163. Interestingly, the largest genes, such as *PTPRT*, *PLCB1* and dJ631M13.5 are located adjacent to or within gene-poor regions.

The repeat content of chromosome 20 is 42%. The distribution of the main classes of repeats (detailed in Supplementary Information) is shown in Fig. 1. Regions of high gene density seem enriched in short interspersed elements (SINEs).

Segmental duplications are another interesting feature of the genome. We compared the masked sequence of chromosome 20 with the rest of the genome and with itself to identify inter- and intrachromosomal duplications, respectively. The segments of chromosome 20 involved in interchromosomal duplications (Fig. 2) often contain pseudogenes; for example, at 6 Mb, AL359954 contains a pseudogene similar to TRDBP that maps to 1p36 (AL109811). The region at 53.9 Mb (Fig. 2) that is duplicated in chromosomes 21 and 22 was recently described as part of a breast cancer amplicon¹⁶. A region of about 500 kb between 25.8 and 26.3 Mb is implicated in both types of segmental duplications. A core region of 100 kb that harbours a copy of exon 7 of the CFTR gene is duplicated on chromosome 20. The second copy is located in AL121762-AL441988 at the pericentromeric region of the q arm. Copies of the extended region seem to be present on chromosomes 9, 12, 15, 17 and 19 (Fig. 2). Secondary signals in the pericentromeric regions of these and other chromosomes were also observed on FISH analysis of clones AL078587 and AL121762. It will be interesting to investigate whether the gene structures annotated in AL121762 and AL441988 are expressed genes, particularly C20orf80, which is similar to the FRG1 gene. FRG1 is located 100 kb centromeric of the repeat units on chromosome 4q35, which are deleted in facioscapulohumeral muscular dystrophy. The region from 48.2 to 48.8 Mb is bordered by two copies of a 60-kb intrachromosomal duplication.

The integrated Marshfield male, female and sex-averaged genetic maps of chromosome 20 (ref. 17) were aligned to the physical map (Fig. 3). The steepest increase in recombination frequency is observed between markers D20S178 and D20S176, which are both located in the region of duplication described above. A

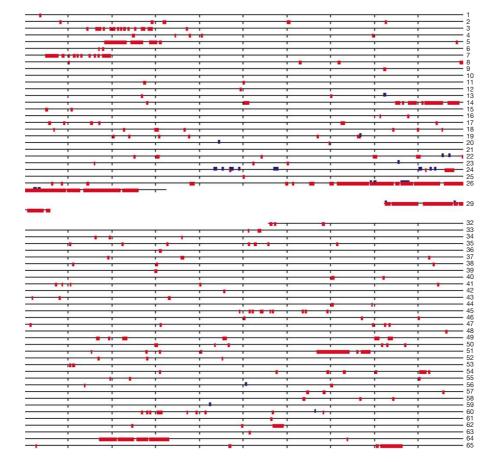


Figure 2 Duplication landscape of chromosome 20. Intrachromosomal and interchromosomal duplications are shown in blue and red, respectively. Each horizontal line represents 1 Mb of the sequence from the telomeric end of the short arm (top left) to the telomeric end of the long arm (bottom right). The gap indicates the centromeric region. Pairwise alignments generated by Exonerate and longer than 1 kb are shown.

region of very low recombination extends for about 20 Mb between markers D20S432 and D20S859. The rate of recombination in specific loci differs between the two sexes. Compared with that of the female, the rate of male recombination is higher along the p arm up to marker D20S432 and lower across the rest of the chromosome (Fig. 3).

Sequence variation

The definition of the common ancestral haplotypes that are present in the population relies on the availability of an extensive collection of single nucleotide polymorphisms (SNPs). We first placed 26,678 SNPs (deposited in the dbSNP database, http://www.ncbi.nlm.nih. gov/SNP) on the sequence of chromosome 20. Of those, 13,016 were derived from sequence analysis of clone overlaps by the program ssahaSNP¹⁸. To recover additional SNPs in clone overlaps, we realigned all available clone-based shotgun sequences from chromosome 20 (including unfinished sequence in clone overlaps that was previously archived and therefore excluded from the earlier analysis) onto the finished sequence with ssahaSNP, and detected 11,050 SNPs (submitted to dbSNP). Merging the two data sets resulted in 32,763 unique SNPs on chromosome 20 (Fig. 1), of which 6,085 are new. In the unique set, there are 14,211 SNPs (43.4%) located within annotated genes and 3,061 of them are in exons.

Comparative analysis

Functional features such as exons and regulatory elements have been conserved through evolution and there is compelling evidence that comparative genomic sequence analysis is a powerful tool in the quest to complete the structural annotation of the human genome. Two data sets were available in the public domain at the time of analysis: about 13 million sequence reads of a mouse whole-genome shotgun giving an estimated genome coverage of 2.3-fold (http:// trace.ensembl.org), released by the Mouse Sequencing Consortium on 8 May 2001; and 816,262 single sequence reads from BAC and plasmid ends of the *T. nigroviridis* genome, totalling 663,839,518 bases and corresponding to 1.72 genome equivalents, generated at Genoscope. Thus we undertook the comparative analysis of the finished and annotated sequence of an entire human chromosome against two vertebrate genomes.

Mouse sequences were aligned to the sequence of chromosome 20 using Exonerate version 0.3d (Guy StC. Slater, unpublished). We obtained matches with 63,644 mouse sequences representing 12,041 regions of sequence conservation (RSC) along chromosome 20. Tetraodon sequences were aligned at Genoscope by Exofish ('exon finding by sequence homology'), which generates ecores (evolutionary conserved regions)¹⁹. Matches were obtained with 2,992 ecores (available at http://www.genoscope.cns.fr/exofish). We first examined the annotated gene structures; 77.4% of the 727 genes and 89% of the 168 pseudogenes have at least one exon matched by a mouse RSC or Tetraodon ecore. This figure is much higher for the 557 genes in groups 1 and 2 (94%) than it is for the 'putative' gene structures in group 4 (33%). Furthermore, the two sets differ in the ratio of genes with only a mouse RSC to genes with both an RSC and ecore match: 1:3.8 and 1:0.2, respectively. These observations suggest that the putative gene structures may represent largely UTRs, which have sequences known to be less well conserved between species, and possibly genes that appeared later in evolution.

We then looked at matches outside annotated exons as a way to assess the completeness of the current annotation. Such matches may correspond to exonic sequences that have not been annotated in the present study owing to lack of supporting evidence (for example, EST, cDNA and protein homologies). Note that we did not use RSCs and ecores during the annotation process. We found 5,447 RSC and 207 ecore matches, and 60 of these non-exonic regions are conserved in all three species. Of all annotated exons (including pseudogenes), 2,050 (36.3%) contain a region conserved in all three

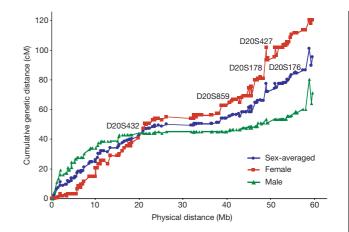


Figure 3 Alignment of the genetic map of chromosome 20 to the physical map. The two maps are aligned from the telomeric end of the short arm to the telomeric end of the long arm. The position of each genetic marker on the female, the male and the sex-averaged genetic map is indicated.

species (in contrast to 0.2% of the annotated introns). Thus, we postulate that about 97.2% (2,050 / (2,050 + 60)) of all coding exons of chromosome 20 have been annotated in this study. A caveat is whether the set of annotated genes used in this analysis is representative of genes that appeared recently in evolution, as ecores are biased to more conserved genes; however, we consider that such an effect cannot be substantial.

The 639 and 4,808 RSC matches in annotated introns and intergenic regions, respectively, suggest that although the mouse data set provides better coverage (70% of all exons) than the ecores, exonic sequences cannot be readily identified by simple comparison at the DNA level.

GENSCAN can be used on small segments of genomic sequence to effectively evaluate the likelihood of that segment containing an exon. We performed a GENSCAN analysis on 'extended RSCs', which included 100 bp of human sequence either side of the RSC match, to divide them into those that were more likely to be coding regions of sequence conservation (cRSC) and those more likely to be noncoding. This predicted 3,299 cRSCs and 8,836 noncoding RSCs and found that 65.7% of cRSCs match annotated exons (3.8% are within introns). As a result, the 874 cRSCs found between annotated genes is a set enriched in regions that may represent nonannotated exons (there are 4,748 RSC matches in intergenic regions).

Conclusion and medical implications

We sequenced the euchromatic portion of human chromosome 20 leaving four small gaps that account for no more than 320 kb. In the 59,421,637 bp of sequence, we annotated 727 gene structures of which 64% are complete and 168 pseudogenes. A comparison of this product with the draft assembly of chromosome 20 reported earlier this year² clearly shows the importance of generating a contiguous finished reference sequence for each human chromosome. Both the G+C and gene density plots of chromosome 20 peak between 60 and 62.5 Mb (Fig. 1) at the qtel region, which is in sharp contrast to the corresponding plots shown in Fig. 11 in ref. 2, which peak at least 12 Mb proximal of the telomere. Furthermore, the order in which genes are shown in the magnified part of Fig. 13 in ref. 2 is incorrect. For example, the gene OSBPL2 (oxysterol binding protein 2) and bB379O24.1 (GATA5 related) are located at 60.2-60.6 Mb and cannot map between PTPRT (protein tyrosine phosphatase, receptor type) at 41 Mb and ZNF217 (Kruppel-like transcription factor) at 51.7 Mb. The use of the clone map information was instrumental in resolving similar problems during the assembly of the chromosome 20 draft sequence.

The output of the comparative analysis of chromosome 20 from the mouse whole-genome shotgun and the ecores generated from the Tetraodon genomic sequence suggests that the current sets of human and mouse ESTs and 'full-length' cDNAs together with the proteomes of model organisms are adequate to allow the identification of the vast majority of human genes in the sequence. As expected, we found that comparative analysis can be used to reliably identify exonic sequences. The mouse shotgun data alone cannot be used reliably to postulate the number of non-annotated exons, owing to the overall higher degree of sequence conservation. The use of the two data sets together, however, provides an excellent tool for assisting the identification of new, and the completion of existing, gene structures. In the present study, the ability to identify regulatory elements in the sequence of chromosome 20 by comparison to the mouse sequence data can be substantiated only by anecdotal evidence. A three-way comparison with the addition of the genome sequence of a species more closely related to humans may hold the key in this endeavour²⁰.

Chromosome 20 is best known for harbouring the genes that cause Creutzfeldt-Jakob disease (PRNP) and severe combined immunodeficiency (ADA). However, the causes of the sporadic cases of Creutzfeldt-Jakob disease (80% of all cases) remain unknown, and no mutation in the ADA gene has been identified to explain the phenotype of ADA excess in haemolytic anaemia. Furthermore, there are still single-gene disorders mapped to chromosome 20 (http://www.ncbi.nlm.nih.gov/Omim) for which the underlying genetic defect is not known. The resources generated by the Human Genome Project have already been used to accelerate the cloning of disease genes on chromosome 20; the Alagille (JAG1)²¹, McKusick-Kaufman (MKKS)²², ICF (DNMT3B)²³ and Hallervorden-Spatz (PANK2)²⁴ syndromes are recent examples. The reported finished and annotated sequence and its variation will be a valuable tool in tackling not only the remaining single-gene diseases but also the multifactorial diseases that have been linked to chromosome 20, such as type 2 diabetes, obesity, cataract, eczema and Grave's disease. Evidence for a susceptibility locus for hereditary prostate cancer on 20q13 has also been reported^{25,26}. In addition to the sequence itself, the isolated clones used in the sequencing process constitute a unique resource in studying chromosome loss and/or amplification in various types of cancer. We have recently reported the refinement of a commonly deleted region (CDR) of 20q12-13.1 found in patients with myeloproliferative disorders and myelodisplastic syndromes²⁷. Others have reported the characterization of a breast cancer amplicon at 20q13.2 (ref. 16), whereas several studies have reported loss of heterozygosity across regions of 20q using comparative genomic hybridization^{28,29}.

Methods

Clone map and sequence assembly

Clone map construction is described in ref. 30. Mapped sequence tagged sites (STSs) for screening genomic PAC and BAC libraries were selected from the integrated radiation hybrid map constructed for chromosome 20 (http://www.sanger.ac.uk/cgi-bin/rhtop? chr=20), which harbours 1,493 STS-based markers. In regions with no clone coverage, screening was extended to the CEPH (Centre d'Etude du Polymorphisme Humain), ICRF (Imperial Cancer Research Fund) and ICI (Imperial Chemical Industries) YAC libraries and the LANL (Los Alamos National Laboratory) chromosome-20-specific cosmid library (links for the libraries can be found at http://www.hgmp.mrc.ac.uk/Biology/descriptions/ genomic_libraries.html). For the shotgun phase, pUC plasmids with inserts of 1.4-2 kb were sequenced from both ends by the dideoxy chain termination method³¹ with big dye terminator chemistry³². Most of the reactions were analysed on ABI3700 capillary sequencing machines. The resulting data were processed by a suite of in-house programs (http://www.sanger.ac.uk/Software/sequencing) before assembly with the PHRED^{33,} ⁵⁴ and PHRAP (http://www.phrap.org) algorithms. For the finishing phase, we used the GAP4 program35 to help assess, edit and select reactions, eliminate ambiguities and close sequence gaps. Sequence gaps were closed by a combination of primer walking, PCR, short/long insert sublibraries36, sublibrary screening with oligonucleotides and, in rare cases, transposon sublibraries.

Sequence analysis tools

Interspersed and simple tandem repeats were identified with Repeatmasker (http:// repeatmasker.genome.washington.edu) and etandem (http://www.emboss.org), respectively. BLAST 1.4 (default parameters and matrix; http://blast.wustl.edu) was used to identify initial matches, which were then re-aligned by EST_GENOME37. BLASTN was used with a 65% similarity cutoff in the comparison against the RIKEN mouse cDNA set³⁸ instead of 95%, which is used when searching human ESTs, to find significant matches. In the unmasked sequence, CpG islands were predicted by searching for sequence segments that are at least 400 bp, have a G+C content greater than 50%, and an expected/observed CpG count of greater than 0.6. The completed analysis was assembled into contigs and visualiszed in AceDB (http://www.acedb.org), whereas an Ensembl (http://www. ensembl.org) database of the sequence assembly and the annotated genes was constructed and used for calculation of statistics and producing Fig. 1. In SNP analysis, only those regions of chromosome 20 where a SNP was detected by at least four reads was considered valid, since the depth of shotgun sequencing for these clones was greater than 4×. The phred-quality value of at least four of the reads at the SNP location had to be at least 30 (error probability of phred base calling 0.001 or less). Exonerate was run with an initial word length of 14 bp, gap penalties of 8 for opening a gap and 4 for extending one, and a score of 5 and -4 for DNA matches and mismatches, respectively.

Received 13 August; accepted 25 October 2001.

- International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 409, 860–921 (2001).
- 2. Venter, C. J. et al. The sequence of the human genome. Science 291, 1304-1351 (2001).
- Riethman, H. C. et al. Integration of telomere sequences with the draft human genome sequence. Nature 409, 948–951 (2001).
- Chute, I., Le, Y., Ashley, T. & Dobson, M. J. The telomere-associated DNA from human chromosome 20p contains a pseudotelomere structure and shares sequences with the subtelomeric regions of 4q and 18p. *Genomics* **41**, 51–60 (1997).
- Felsenfeld, A., Peterson, J., Schloss, J. & Guyer, M. Assessing the quality of the DNA sequence from the Human Genome Project. *Genome Res.* 9, 1–4 (1999).
- Morton, N. E. Parameters of the human genome. *Proc. Natl Acad. Sci. USA* 88, 7474–7476 (1991).
 Altschul, S. F., Gish, W., Miller, W., Meyers, E. W. & Lipman, D. J. Basic local alignment search tool.
- Mol. Biol. 215, 403–410 (1990).
 Salamov, A. A. & Solovyev, V. V. Ab initio gene finding in Drosophila genomic DNA. Genome Res. 10,
- 51 Salantov, A. A. & Solovyev, V. *Ab initia* gene initiang in *Dissiphila* genomic DNA. *Genome* Res. 10, 516–522 (2000).
- Burge, C. & Karlin, S. Prediction of gene structures in human genomic DNA. J. Mol. Biol. 268, 78–94 (1997).
- Hattori, M. et al. The DNA sequence of human chromosome 21. The chromosome 21 mapping and sequencing consortium. Nature 405, 311–319 (2000).
- 11. Dunham, I. et al. The DNA sequence of human chromosome 22. Nature 402, 489-495 (1999).
- 12. Deloukas, P. et al. A physical map of 30,000 human genes. Science 282, 744-746 (1998).
- Antequera, F. & Bird, A. Number of CpG islands and genes in human and mouse. Proc. Natl Acad. Sci. USA 90, 11995–11999 (1993).
- Hayward, B. E. & Bonthron, D. T. An imprinted antisense transcript at the human GNAS1 locus. Hum. Mol. Genet. 9, 835–841 (2000).
- Brett, D. et al. EST comparison indicates 38% of human mRNAs contain possible alternative splice forms. FEBS Lett. 474, 83–86 (2000).
- Collins, C. *et al.* Comprehensive genome sequence analysis of a breast cancer amplicon. *Genome Res.* 11, 1034–1042 (2001).
- Yu, A. et al. Comparison of human genetic and sequence-based physical maps. Nature 409, 951–953 (2001).
- Ning, Z., Cox, A. J. & Mullikin, J. C. SSAHA: A fast search method for large DNA databases. *Genome Res.* 11, 1725–1729 (2001).
- Roest Crollius, H. et al. Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence. Nature Genet. 25, 235–238 (2000).
- Dubchak, I. et al. Active conservation of noncoding sequences revealed by three-way species comparisons. Genome Res. 10, 1304–1306 (2000).
- Li, L. et al. Allagile syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1. Nature Genet. 16, 243–251 (1997).
- 22. Stone, D. L. *et al.* Mutation of a gene encoding a putative chaperonin causes McKusick-Kaufman syndrome. *Nature Genet.* **25**, 79–82 (2000).
- Hansen, R. S. et al. The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc. Natl Acad. Sci. USA 96, 14412–14417 (1999).
- Zhou, B. et al. A novel pantothenate kinase gene (PANKS) is defective in Hallervorden-Spatz syndrome. Nature Genet. 28, 345–349 (2001).
- Berry, R. et al. Evidence for a prostate cancer-susceptibility locus on chromosome 20. Am. J. Hum. Genet. 67, 82–91 (2000).
- Zheng, S. L. et al. Evidence for a prostate cancer linkage to chromosome 20 in 159 hereditary prostate cancer families. Hum. Genet. 108, 430–435 (2001).
- Bench, A. et al. Chromosome 20 deletions in myeloid malignancies: reduction of the common deleted region, generation of a PAC/BAC contig and identification of candidate genes. Oncogene 19, 3902– 3913 (2000).
- Bench, A. *et al.* A detailed physical and transcriptional map of the region of chromosome 20 that is deleted in myeloproliferative disorders and refinement of the common deleted region. *Genomics* 49, 351–362 (1998).
- Rigaud, G. *et al.* High resolution allelotype of nonfunctional pancreatic endocrine tumors: Identification of two molecular subgroups with clinical implications. *Cancer Res.* 61, 285–292 (2001).
- Bentley, D. R. *et al.* The physical maps for sequencing human chromosomes 1, 6, 9, 10, 13, 20 and X. *Nature* 409, 942–943 (2001).
- Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. Proc. Natl Acad. Sci. USA 74, 5463–5467 (1977).
- Rosenblum, B. B. et al. New dye-labeled terminators for improved DNA sequencing patterns. Nucleic Acids Res. 25, 4500–4504 (1997).
- Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8, 175–185 (1998).

Second Se

- Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8, 186–194 (1998).
- Bonfield, J. K., Smith, K. F. & Staden, R. A new DNA sequence assembly program. Nucleic Acids Res. 23, 4992–4999 (1995).
- McMurray, A. A., Sulston, J. E. & Quail, M. A. Short-insert libraries as a method of problem solving in genome sequencing. *Genome Res.* 8, 562–566 (1998).
- Mott, R. EST_GENOME: a program to align spliced DNA sequences to unspliced genomic DNA. Comput. Appl. Biosci. 13, 477–478 (1997).
- The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection. *Nature* 409, 685–690 (2001).

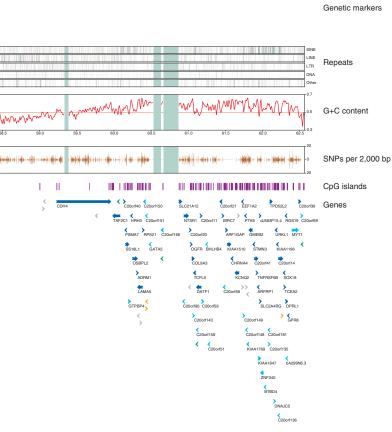
Supplementary Information accompanies the paper on *Nature*'s website (http://www.nature.com).

Acknowledgements

All the reported DNA sequences have been deposited to the EMBL, GenBank or DDBJ databases. We thank H. Roest Crollius, O. Jaillon and the *Tetraodon* group at Genoscope for the Exofish data; E. Zdobnov for the InterPro analysis; the Ensembl team for assistance in the computational analysis; and the group of M. S. Jackson at the University of Newcastle upon Tyne for markers and clones that assisted the mapping of the pericentromeric duplication. Finally, we thank H. Wain, E. Bruford, M. Lush, R. Lovering, M. Wright, P. Shetty and S. Povey for help with gene nomenclature and the Wellcome Trust for financial support.

Correspondence and requests for material should be addressed to P.D. (e-mail: panos@sanger.ac.uk).

Contig tiling path	AL121758 AL031665 AL AL049761 AL050225 AL0496 AL0354548 AL133231 AL161929 AL138531 AL AL360078 AL121747 AL118502 AL110114 AL109568	AL109804 AL309016 AL040712 AL121905 AL127773 AL121875 9694 AL21707 AL221875 AL30916 AL30916 AL309172 AL121905 AL309173 AL121875 9694 AL12780 AL204805 AL00976 AL309876 AL309176 AL309176	AL121890 AL38986 AL035461 AL192504 AL33354 AL121755 AL035349 AL106918 AL AL22096 AL09641 AL121757 AL17277 AL17805 AL129606 AL036664 J21916 AL445075 AL121904 AL109835 AL390028 AL396964 AL121911 AL	AL031655 A021396 AL03455 AL03455 AL021498 234554 AL080248 AL09705 AL021879 AL021406 AL050319 AL121898 AL151655 AL09658 AL031653 AL035664 AL121599 AL048662 AL44567 AL02365 AL151655 AL09679 AL026778 AL031679 AL050323 AL365212 AL031683 AL121909 AL	ALGS1482 AL 133340 ALO46649 ALG364 AL109754 ALG34430 ALG56433 ALG7858 AL109827 AL680274 135953 AL353739 ALG25913 ALG5566 AL16842 ALG78337 AL45667 S3612 ALG3427 AL354824 AL355596 AL135537 AL65680 AL161588 A	48 AL078623 AL000020 AL121754 AL117333 AL139440 AL139893 AL159098 AL159098 AL159088 AL AL19848 AL034451 AL133331 AL133453 AL198657 AL1399 AL034547 AL066882 AL55480 AL45589 AL121782 AL161689 AL35378	AL079538 AL354773 AL01677 AL121884 A000198 AL161941 AL155938 AL01184 A135828 AL13828 AL38808 AL049677 AL05573 AL357504 AL04974 AL121892 AL121892 AL151800 AL049633 AL050224 AL12091 AL116509 AL03429 AL118 AL355133 AL118510 AL133462 AL15008 AL118503 AL357654 AL17276 AL355579 AL12192 A	AL121779 AL049546 AL03075 AL03045 AL00021 AL121893 AL035563 Seko AL031664 AL160071 AL16411 AL156176 AL121780 AL139598 AL139511 AL122765 AL121585 AL389697 AL121800 AL135536 A	AL121830 AL161956 AL09425 AL09558 AL121721 AL07854 AL1. 200 AL139827 AL121761 AL09549 AL15089 AL150963 AL09647 AL138429 AL132821 AL09648 AL121866 AL13778 /	AL121912 AL130225 AL117875 AL36733 AL198175 AL3665 AL117332 AL121821 AL030558 AL449003 AL170120 AL121390 AL157412 AL133464 AL04973 AL121739 AL158013 AL03562 AL09607 AL049737 AL121722 AL3	ALS91074 AL119503 AL157413 AL0803 048651 AL390037 AL122884 AL121851 AL09426 AL355582 AL11808 AL098677 AL35453 AL133466 AL158096 AL121925 AL0 132 AL357508 AL109554 AL356967 AL109830 AL45854 AL035661	512 <u>AL355812 AL031673 AL391119</u> 035522 AL031673 AL391119 AL121772 AL191502 AL073587 AL39116	Contig tiling path
Genetic markers	ມລະອີ້າ155 ນະດີຮ້າຍ3 ນະດີຮ້າຍ9 ນະວີຮ້ອຍ6 ນະລີ້້າ157 ນະດີຮ້າຍ5 ນະລີ້ຮາຍ4 ນະລີ້ຮ່າ17 ນະລີ້ຮ້າຍ5	ວຣ ວລຣີສາອ ວລຣີສາອ ລະເອົາລັດ ເຊິ່ງ ເຊິ່ງ ເຊິ່ງ ເຊິ່ ເຊິ່ງ ເຊິ່ງ ເຊິ່ ເຊິ່ງ ເຊິ່ງ ເຊິ່	සේ සංවේශාෂාවයාගීමයක සංවේශාෂාව සංවේශාෂාව සංවේශාෂාවය මෙත සංවේශාෂාව සංවේශාෂාව සංවේශාෂාව සංවේශාෂාව මේ සංවේශාෂාව සංවේ සංවේශාෂාව සංවේශාෂාව සංවේෂාව සංවේශාෂාව සංවේ සංවේශාෂාව සංවේශාෂාව සංවේශාෂාව සංවේශාෂාව සංවේශාෂාව සංවේ සංවේශාෂාව සංවේශාෂාව සංවේශාෂාව සංවේශාෂාව සංවේශාෂාව සංවේ සංවේශාෂාව සංවේශාෂාව සංවේශාෂාව සංවේශාෂාව සංවේශාෂාව සංවේශා සංවේශාෂාව සංවේශාෂාව සංවේශාෂාව සංවේශාෂාව සංවේශාව සංවේශා සංවේශාෂාව සංවේශාෂාව සංවේශාව සංවේශාව සංවේශාව සංවේශාව සංවේශාව සංවේ සංවේශාව සංවේශාව සංවේශාව සංවේශාව සංවේශාව සංවේශාව සංවේ සංවේශාව සංවේශාව සංවේශාව සංවේශාව සංවේශාව සංවේශාව සංවේ සංවේ සංවේශාව සංවේශාව සංවේශාව සංවේශාව සංවේශාව සංවේ සංවේශාව සංවේ සංවේශාව සංවේශාව සංවේශාව සංවේශාව සංවේශාව සංවේ සංවේ සංවේශාව සංවේශාව සංවේශාව සංවේශාව සංවේ සංවේශාව සංවේශාව සංවේශාව සංවේශාව සංවේශාව සංවේ සංවේ සංවේශාව සංවේශාව සංවේ සංවේ සංවේ සංවේ සංවේ සංවේ සංවේශාව සංවේශාව සංවේශාව සංවේ සංවේ සංවේශාව සංවේ සංවේ සංවේශාව සංවේශාව සංවේ සංවේ සංවේ සංවේ සංවේ සංවේ සංවේ සංවේ සංවේ සංවේ සංවේ සංවේ සංවේ සංවේ	වැඩිසිංච ව වැඩිසිංච වැඩිසිංච වැඩ පොඩ්සිංච වැඩිසිංච වැඩ පොඩ්සිංච වැඩිසිංච වැඩ පොඩ්සිංච වැඩිසිංච වැඩ පොඩ්සිංච වැඩිසිංච වැඩ පොඩ්සිංච වැඩිසිංච වැඩ පොඩ්සිංච වැඩිසිංච වැඩ පොඩ්සිංච වැඩිසිංච වැඩ පොඩ්සිංච වැඩිසිංච වැඩ පොඩ්සිංච වැඩිසිංච වැඩිසිංච වැඩිසිංච වැඩිසිංච වැඩිසිංච වැඩිසිංච ව පොඩ්සිංච වැඩිසිංච වැඩිසිංච වැඩිසිංච වැඩිසිංච වැඩිසිංච වැඩිසිංච වැඩිසිංච වැඩිසිංච වැඩිසිංච වැඩිස	ວຍລື້ອາຍວ ວຍລື້ອາຍວ ວຍລື້ອາຍວ ວຍລື້ອາຍວ ວຍລື້ອາຍວ ວຍລື້ອາຍວ ວຍລື້ອຍຈາ ວຍລື້ອຍຈາ ວຍລື້ອວາ ວຍລື້ອຈ	66 Dadënes Dadëens 87	ດເຊື້ອຍຄວລຕີອອອ ເວລີເອົາ ເປັນ ເວລີເອົາເຊ ເວລີເອື້ອກາຣ ເວລາຕີອອຣ ເວລາຕີອອຣ	ວລັ້ອີ114 Daວີ້ອີ182 Daວີ້ອີ475 Daວີ້ອີ112 Daວີ້ອິ480 Daວີ້ອີ431 Daວີ້ອີ470	อะซิษาา อะซิธร อะซิษาง2	น องไร้เรอ องไร้สงช องไร้นา7 คะชั้งเงง	ວວວີ້ຮາຍ4 ວວດີ້ຮາຍ ວວດີ້ຮອຍ ວວດີ້ຮອບ ວວດີ້ຮາດາ	5 D205191	Genetic markers
Repeats	SINE												Sine LINE LITE Repeats DNA Other
G+C content	0.7 0.5 0.3 0.5 1.0 1.5	5 20 25 30 35 40 45	s so ss éo és	7:0 75 80 85 80 85	160 rés ris ris ris	20 125 130 135 140	145 150 155 160 165 170	17.5 160 165 150	145 200 205 2	M. M	14 245 240 245 250	MMMM Mun zás zás zás zás	0.7 -0.5 G+C content
SNPs per 2,000 bp	20 0	- 	essen <mark>deke ene f∳el</mark> sere > ternenik tribeke <mark>lskytekenik enerskenikenik</mark> e	n fan ar Hennen (1997) (1997) (1997) (1997) (1997) - Nene Henne (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997) (1997)	tall from the finite of the second	ar - 1994 alle ad feifine i fei <mark>nd alle de alle de la addeana</mark> e an	# -1 488 0 +111 + 14	n41948) = 103/1 18 − 4 − 30/13) \$ \$ 30 × 10 × 10 × 100 \$ \$\$\$\$\$\$400 × 104 [>= 10+4+10 \$ (= 10+10+10+10+10+10+10+10+10+10+10+10+10+1	ondialoushi boodaadha oo istata ida baadata ah oo oo oo oo ahaa ahaa	n - een da maken areken kalden kalden hereken kalden kalden kalden hereken kalden kalden kalden hereken hereken	♦∰(\$):\$\$ \$	Re 🗍 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 -	• SNPs per 2,000 bp
CpG islands Genes Known Novel CDS Novel transcript Pseudogene Putative	COURS SOURCE SO	International Control Control	Image: Second		I I II II I II I II I II II I Aurops Marcis Augus Augus Coonter Coonter The subjects Coonter Troa Subjects Coonter Subjects Coonter S	s (*) Caurta Caurta Caurta and Caurta Caurta Caurta and Caurta Caurta Caurta caurta Caurta Caurta Caurta caurta Caurta Caurt	C200rf66 🮸 《 C200rf23 SNRPB2 3	PCSK2 DaTN SNXS CSRP28P POLRSF C20uf78 SL BFSP1 C20uf719 ZNF339 ZNF733 C20uf78 SL Y RBP1 C20uf72 C20uf72 C20uf78 SEC238 C20uf78 SEC238 C20uf84 R88P48 SEC238 C20uf84 SEC238 C20uf84	`	CZDorfie) NIOZB RPL41L2 W HNF38 (CZDorfie) NIOZB RPL41L2 W HNF38 (CZDorfie) NIOZB PAX1 CZDorfie (NIOZD (NIOZD (NIOZD (CZDorfie) NIOZB (CZDorfie) NIOZB (CZDorfie) NIOZB (CZDorfie) NIOZB (NIOZB (CZDorfie) NIOZB (CZDO	STR4 NXT1 CST2 CST2 C20xH3 CST7 THED 2xH336 CST4 CST5 C20xH3 C10R1 MAPB CST1 GGTLA3 V5X1 CST1 CST2 CST1 POM121L3	ENTPOR RUMANNO COUNTING PYGB ZUIF337 C200rf81 ZEX0rf22 C200rf147 C200rf81 KUAG186 CCC LeST2	CpG islands Genes
				AL080249									
Contig tiling path	AL 441998 AL 121702	AL09533 AL095333 AL09533 AL09533 AL095333 AL09533 AL09533	AL109827 AL109502 AL305622 AL121506 AL07960 AL305623 AL121753 AL309675 AL10965 AL09766 AL15173 AL09705 AL13044 AL309674 AL36624 AL121 See AL054568 AL109923 AL109624 AL339825 AL1212752 AL357374 AL36888 BL005458 AL109923 AL109624 AL339825 AL121752 AL357374 AL36888 BL005458 AL109923 AL109624 AL39858 BL005458 AL109624 AL39858 BL005458 BL0	ALISOS16 ALISOS16 ALISOS16 ALISOS16 ALISOS16 ALISOS64 ALISOS66 ALISOS666 ALISOS666 ALISOS666 ALISOS66 ALISOS666 ALISOS666 ALISOS666 ALISOS666 ALISOS66 ALISOS66 ALISOS66 ALISO	2003 AL118523 AL06691 AL01 AL0305251 AL109611 AL2121828 AL035695 AL0160 AL02122 AL0305251 AL109611 AL2121828 AL035695 AL031606 AL02122 AL0306061 AL139163 AL0369050 AL13142 AL03431 AL030522 AL13 AL89 AL121588 AL122181 AL031226 AL03452 AL02017 AL022344 AL021543 AL022643 AL03452 AL030177 AL022344 AL121588 AL122181 AL031206 AL03452 AL020317 AL022344 AL021543 AL022647 D205805 D205855 AL031452 AL032345	A1227763 A103766 AL03166 AL03166 AL03166 57 AL03166 AL03166 AL03166 58 AL03166 AL03166 AL03166 605 AL03166 AL03166 AL03166 57 AL03166 AL03166 AL03166 50 AL03166 AL03166 AL03166 60 AL04672 Z0262 AL03166 AL03167 AL221674 AL046872 Z02239 AL13646 AL171734 AL13866 2058170 D05856 D056589 D056589 D056589 D056589	AL121587 AL17382 Z5016 AL10955 AL131227 188 AL33777 AL335442 AL19952 AL10959 AL021576 AL31677 AL102450 AL45286 AL305089 AL031577 AL305442 AL199322 AL00825 AL008266 AL217778 AL02826 AL31167 AL102450 AL161944 AL30419 AL32777 AL118622 AL047767 AL001826 AL001826 </td <td>034424 AL031666 AL354796 ALM9540 AL03418 AL139051 AL137078 AL135050 AL121778 AL90212 AL021394 226330 AL133341 A 54745 AL38950 AL30211 AL353777 AL131777 AL130102 AL045441 B88 D205891205197 D205178 AL72188 AL1 S888 D205891205197 D205178 D2051</td> <td>ALDIS7685 ALDIS768 ALDIS768 ALT02185 ALT02185 ALDIS423 A 541 ALT3024 ALT3074 ALT02563 ALT0772 AL354 445192 AL095437 AL357560 ALT0523 AL031885 AL101937 Set27 D205445 D2</td> <td>AL024553 AL025457 AL025684 AL395554 AL1217 AL024429 AL02469 AL1217155 AL12868 AL02429 AL12177 B AL50563 AL025443 AL029339 AL03579 AL10984 AL13329 AL13828 AL121915 AL036882 AL38807 AL035072 AL02 AL035679 D26899 D26857 D26845 D2684</td> <td>ALDER765 ACCOSCO AL396983 3 AL121902 AL364772 ACCOSCO AL396931 AL139335 AL A193910 AL161945 ACCOS655 AL15783 ACOUT762 AL39835 AL1622 LI09977 AL051674 AL391697 AL394983 ACCOS680 ACCOS600 AL1622 LI09977 AL051674 AL696316 AL109898 ACCOS680 ACCOS600 AL121212 D265167 D265867 D2658165 D265120</td> <td>AL09867 AL109829 118501 AL121693 AL271865 AL17383 202 AL162294 AL108101 AL199828 AL123914 1310 AL139809 AL055683 AL355686 AL355684 AL359634 22 AL33420 AL158015 AL160413 AL171591 AL139844 AL1 Do05499 Do05853 Do05100</td> <td>L137162 AL035541 AL356984 AL109840 AL122822 AL122222 AL135599 AL121913 AL035455 AL135348 AL035250 AL121968 AL109638 AL121200 AL127308 AL122558 AL16176 AL162291 AL356854 AL035037 AL121917 AL121919 AL589889 AL158082 AL035070 A 8806 AL161657 AL17714 AL19955 AL161943 AL356776 AL118513 AL12855 AL19532 AL160410 AL207503 AL35577 AL121 D205168 D205430 D05164 D05544 D205171</td>	034424 AL031666 AL354796 ALM9540 AL03418 AL139051 AL137078 AL135050 AL121778 AL90212 AL021394 226330 AL133341 A 54745 AL38950 AL30211 AL353777 AL131777 AL130102 AL045441 B88 D205891205197 D205178 AL72188 AL1 S888 D205891205197 D205178 D2051	ALDIS7685 ALDIS768 ALDIS768 ALT02185 ALT02185 ALDIS423 A 541 ALT3024 ALT3074 ALT02563 ALT0772 AL354 445192 AL095437 AL357560 ALT0523 AL031885 AL101937 Set27 D205445 D2	AL024553 AL025457 AL025684 AL395554 AL1217 AL024429 AL02469 AL1217155 AL12868 AL02429 AL12177 B AL50563 AL025443 AL029339 AL03579 AL10984 AL13329 AL13828 AL121915 AL036882 AL38807 AL035072 AL02 AL035679 D26899 D26857 D26845 D2684	ALDER765 ACCOSCO AL396983 3 AL121902 AL364772 ACCOSCO AL396931 AL139335 AL A193910 AL161945 ACCOS655 AL15783 ACOUT762 AL39835 AL1622 LI09977 AL051674 AL391697 AL394983 ACCOS680 ACCOS600 AL1622 LI09977 AL051674 AL696316 AL109898 ACCOS680 ACCOS600 AL121212 D265167 D265867 D2658165 D265120	AL09867 AL109829 118501 AL121693 AL271865 AL17383 202 AL162294 AL108101 AL199828 AL123914 1310 AL139809 AL055683 AL355686 AL355684 AL359634 22 AL33420 AL158015 AL160413 AL171591 AL139844 AL1 Do05499 Do05853 Do05100	L137162 AL035541 AL356984 AL109840 AL122822 AL122222 AL135599 AL121913 AL035455 AL135348 AL035250 AL121968 AL109638 AL121200 AL127308 AL122558 AL16176 AL162291 AL356854 AL035037 AL121917 AL121919 AL589889 AL158082 AL035070 A 8806 AL161657 AL17714 AL19955 AL161943 AL356776 AL118513 AL12855 AL19532 AL160410 AL207503 AL35577 AL121 D205168 D205430 D05164 D05544 D205171
Contig tiling path Genetic markers	AL 441988 AL 1227782	AL049599 AL034500 AL199206 AL024500 AL199206 AL024500 AL12751 AL100175 AL330926 AL121568 AL058071 AL197226 AL054621 AL054600 AL17381 AL051658 AL121680 AL121050 AL121001 AL19206 AL121720 AL170115 AL359766 AL121987 AL135843 AL390298 AL121901 AL19205 AL054 D205464 D205111 D205874 D205903 D205843 D205465 D205843 D205845 D205845 D205845	AL108627 AL108626 AL27568 AL27568 AL27568 AL27568 AL27568 AL27568 AL27568 AL27568 AL27568 AL27575 AL20965 AL20065 AL20965 AL20	885 AL380374 AL132786 AL391114 AL380014 AL118499 AL031651 AL391653 AL03165 AL512784 AL031662 AL365605 AL034422 AL182293 AL109823 AL39655 AL035419 AL12	2003 A118223 ALON251 ALO AL030251 AL106611 AL121268 AL030665 AL03160 AL03026 AL30850 AL011611 AL121268 AL030665 AL03160 AL03026 AL13 AL30850 AL12158 AL123419 AL031256 AL03452 AL03652 AL03452 AL03452 AL03452 AL03452 AL03452 AL03453 AL03537 AL035374 AL025452 AL03453 AL035355 AL035555 AL035555 AL035555 AL035555 AL035555 AL035555 AL035555 AL035555 AL035555	<u>А.127793</u> 10197 А.13229 <u>А.03196</u> <u>А.03196</u> <u>А.03196</u> 57 А.031909 <u>А.03197</u> А.103969 <u>А.135999</u> <u>А.03197</u> А.202198 2025 А.1524912 А.049912 203942 <u>А.109828</u> <u>А.031978 А.202198</u> <u>201973</u> <u>А.121874 А.139994 А.03473 А.022239</u> <u>А.138461</u> <u>А.117374</u> <u>А.138800</u> <u>D.005466</u> <u>D.005466</u>	AL121587 AL117382 289016 AL10856 AL121587 AL131527 AL21587 AL1075462 AL138528 AL10869 AL021576 AL01671 AL162458 AL445086 AL055688 AL035417 257053 AL00872 AL05876 AL051671 AL162458 AL451687 AL3 AL505688 AL035417 257053 AL00872 AL05876 AL05167 AL3 AL111544 AL034119 AL132772 AL118522 AL049767 AL011693 AL055688 AL035682 AL03682 AL03682 AL036882 AL036882 AL036888 AL036888 AL03688	201424 L031666 AL355766 AL069640 AL009418 AL139353 AL137078 AL35550 AL357778 AL390212 AL61394 200330 AL133341 A 201555 AL39671 AL39777 AL121777 AL139102 AL09454 86 AL031055 AL39641 AL109825 AL394813 AL37558 AL121888 AL4 8688 D268281D265197 D265178 D2651 Base D268281D265197 D265178 D2651 D	AL039685 AL039766 AL121950 AL121950 AL122170 AL03423 A 541 AL123150 AL121950 AL122170 AL03428 A 245192 AL09537 AL35750 AL19525 AL03185 AL191937 B127 Dx85445 D2 Dx8542		AL067705 AL201620 AL201693 AL201620 AL201693 AL2	AL05667 AL109629 11950 AL12199 AL121968 AL171283 222 AL162294 AL16101 AL109828 AL121941 131 AL19899 AL055683 AL355388 AL353640 AL39554 22 AL133420 AL158015 AL160413 AL121991 AL19894 AL1 D205849 D205843 D205100 D2058492	4.137162 AL039541 AL35484 AL109840 AL13282 AL13022 AL13599 AL121913 AL039545 AL199349 AL030505 AL121956 AL109628 AL13022 AL13599 AL121913 AL039545 AL19934 AL000027 AL21917 AL21919 AL38888 AL18982 AL03070 A AL97820 AL17180 AL12258 AL161943 AL16291 AL354834 AL000027 AL21917 AL21919 AL38888 AL189822 AL03070 A AL97820 AL17180 AL12258 AL161943 AL354776 AL18513 AL132855 AL19632 AL160410 AL357803 AL3535877 AL121 B806 AL16167 AL127414 AL10985 AL161943 AL354776 AL18513 AL132855 AL19632 AL160410 AL357803 AL3535877 AL121 D208158 D208400 D208164 D008171
	AL 41988 AL 121782	D205-425 D205187 D205863 D205843	А.19622 А.19626 А.17366 А.17376 А.19665 А.19662 А.17376 А.19675 А.1965 А.19662 А.17375 А.19675 А.1965 А.19668 А.19773 А.06970 А.19844 А.19864 А.19862 А.197 500 А.05568 А.19923 А.19824 А.1982 А.19724 А.19724 А.196734 А.19863 А.19724 А.19673 А.19823 А.19724 А.19844 А.19864 О.05866 D.05874 D.05866 D.05874 D.05866 D.05874	885 AL380374 AL132786 AL391114 AL380014 AL118499 AL031651 AL391653 AL03165 AL512784 AL031662 AL365605 AL034422 AL182293 AL109823 AL39655 AL035419 AL12		AL121793 AL13229 AL03166 AL03167 AL13229 AL03168 AL03566 AL3695 AL1 AL03168 AL03666 AL36955 AL1 AL12167 AL32934 AL04912 23942 AL09628 AL03167 AL02136 Z4973 AL12167 AL39894 AL04912 23942 AL10982 AL03167 AL02136 Z4973 AL12167 AL39894 AL04913 Z4984 AL02239 AL13846 AL171734 AL13889 D0\$170 D0\$2546 D0\$2546 D0\$2546	AL 121587 AL 117382 28016 AL 106956 AL 13227 186 A333797 AL058482 AL39852 AL09836 AL021576 AL051671 AL16458 AL45382 AL050508 AL05447 20703 AL00872 AL08672 AL01167 AL3 AL011944 AL03419 AL152772 AL11852 AL08776 AL01163 AL00348 AL05588 D26 AL05547 AL01146 D26 D26 100 100 100 100 100 100 100 10	003424 A.0331666 A.135300 A.137778 A.130291 A.033418 A.130293 A.137778 A.133300 A.137778 A.39021 A.037737 A.12777 A.133141 A 85474 A.139670 A.130211 A.037777 A.12777 A.13100 A.L0684 86 A.031055 A.39434 A.109825 A.184813 A.1357568 A.178988 A.4 88 Base D.008991 D.08197 D.008178 D.2084 D.009990 D.00991 D.08178 D.008178 D.2084	4,00966 4,00976 4,00976 4,12160 4,12160 4,12176 4,12015 4,12015 4,1	D205897 D205196 D205185 D2051083 D		4_035667 A_110829 118507 A_1121893 A_1271826 A_1171833 2022 A_1162809 A_035688 A_1535680 A_152954 131 A_138690 A_035680 A_1535688 A_1535680 A_155954 22 A_133420 A_155015 A_160413 A_1271931 A_139834 A_17 D005499 D005883 D005100 D005892	4.117162 AL005541 AL354884 AL109840 A.112802 AL125222 AL135599 AL1271913 AL005455 AL139549 AL005250 AL121908 AL169528 AL271520 AL172808 AL12058 AL180176 AL182291 AL354834 AL005027 AL121917 AL21919 AL389889 AL169502 AL025070 A 8860 AL161657 AL12744 AL109855 AL161943 AL354776 AL18513 AL132655 AL38532 AL160410 AL377503 AL3535877 AL221 D2039158 D2039500 D2039164 D005443 D2039171
Genetic markers Repeats		205425 2051187 205843 205843 205872 205887	Lisses Alisses Alisses <t< td=""><td>895 д.380074 д.127788 д.391114 д.129001 д.118699 д.001651 д.1291680 д.001657 д.1512784 д.0191682 д.385505 д.001462 д.162539 д.1039555 д.0015419 д.12 D205841 D205841 D205841 D205841 D2058478</td><td>artileihos 2285107</td><td>1167 А.132223 А.01160 А.03180 17 А.03160 А.03160 А.03180 А.03180 17 А.03160 А.03160 А.03180 А.03180 17 А.03160 А.03160 А.03180 А.13060 12001 А.03160 А.03160 А.13060 А.13060 12001 1000 Α.03160 Α.131774 Α.13000 12001 1000 Α.03260 Δ.031670 Α.03060 12001 1000 Α.03060 Δ.031670 Α.13060 12001 1000 1000 1000 1000 12001 1000 1000 1000 1000 12001 1000 1000 1000 1000 1000 12001 1000 1000 1000 1000 1000 1000 12001 1000 1000 1000 1000 1000 1000 1000</td><td>AL121597 AL17392 20016 AL00956 AL108227 Bit AL320177 AL00842 AL32922 AL00856 AL21576 AL001671 AL168258 AL462056 AL020076 AL039412 AL00872 AL00877 AL00177 AL00872 <td< td=""><td>201424 L031666 А.13524706 А.109480 А.103418 А.139253 А.121777 А.139253 А.121788 А.139535 А.121788 А.139535 А.121788 А.139535 А.121888 А.1218888 А.1218888</td><td>2005176 2005eee 2005423</td><td>ນຊີ້ສອງ ນຊີ້ສາອ ນະດີ້ສາຍ ນະດີ</td><td>1833 Dafaso Dafa</td><td><u>А.03567</u> <u>А.11950</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u></td><td></td></td<></td></t<>	895 д.380074 д.127788 д.391114 д.129001 д.118699 д.001651 д.1291680 д.001657 д.1512784 д.0191682 д.385505 д.001462 д.162539 д.1039555 д.0015419 д.12 D205841 D205841 D205841 D205841 D2058478	artileihos 2285107	1167 А.132223 А.01160 А.03180 17 А.03160 А.03160 А.03180 А.03180 17 А.03160 А.03160 А.03180 А.03180 17 А.03160 А.03160 А.03180 А.13060 12001 А.03160 А.03160 А.13060 А.13060 12001 1000 Α.03160 Α.131774 Α.13000 12001 1000 Α.03260 Δ.031670 Α.03060 12001 1000 Α.03060 Δ.031670 Α.13060 12001 1000 1000 1000 1000 12001 1000 1000 1000 1000 12001 1000 1000 1000 1000 1000 12001 1000 1000 1000 1000 1000 1000 12001 1000 1000 1000 1000 1000 1000 1000	AL121597 AL17392 20016 AL00956 AL108227 Bit AL320177 AL00842 AL32922 AL00856 AL21576 AL001671 AL168258 AL462056 AL020076 AL039412 AL00872 AL00877 AL00177 AL00872 AL00872 <td< td=""><td>201424 L031666 А.13524706 А.109480 А.103418 А.139253 А.121777 А.139253 А.121788 А.139535 А.121788 А.139535 А.121788 А.139535 А.121888 А.1218888 А.1218888</td><td>2005176 2005eee 2005423</td><td>ນຊີ້ສອງ ນຊີ້ສາອ ນະດີ້ສາຍ ນະດີ</td><td>1833 Dafaso Dafa</td><td><u>А.03567</u> <u>А.11950</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u></td><td></td></td<>	201424 L031666 А.13524706 А.109480 А.103418 А.139253 А.121777 А.139253 А.121788 А.139535 А.121788 А.139535 А.121788 А.139535 А.121888 А.1218888 А.1218888	2005176 2005eee 2005423	ນຊີ້ສອງ ນຊີ້ສາອ ນະດີ້ສາຍ ນະດີ	1833 Dafaso Dafa	<u>А.03567</u> <u>А.11950</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u> <u>А.12193</u>	
	SINE LINE LINE DNA OBJ 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5	008405 D08897 D08892 D08897	.0 32.5 33.0 33.5 34.0	995 <u>а.130074 а.13778</u> <u>а.1391114</u> <u>а.130014 а.119499</u> <u>а.1031651</u> <u>а.139555</u> <u>а.1031657</u> <u>а.1512784</u> <u>а.131652 а.136555</u> <u>а.103452</u> <u>а.142259</u> <u>а.109623</u> <u>а.136555</u> <u>а.103419</u> <u>а.12</u> <u>р.208844</u> <u>р.208899</u> <u>р.208891</u> <u>р.20841</u> <u>р.208479</u> <u>в.20841</u> <u>р.208479</u>	алтланов роблот	9.5 40.0 40.5 41.0 41.5	Dodiest	45.0 45.5 46.0 46.5	Dalassa Dalass D	Dzőšer Dzőřer Dzőřer Dzőřer	48833 D08840 D26877 D268211	53.0 53.5 54.0 54.5	



 ALISTING
 ALISTING

AL590548 5 AL355803 AL158091 AL121581 Contig tiling path