

SUPPLEMENTARY DATA

GBshape: a genome browser database for DNA shape annotations

Tsu-Pei Chiu^{1,#}, Lin Yang^{1,#}, Tianyin Zhou¹, Bradley J. Main¹, Stephen C.J. Parker²,
Sergey V. Nuzhdin¹, Thomas D. Tullius³, and Remo Rohs^{1,4,*}

¹ Molecular and Computational Biology Program, Department of Biological Sciences,
University of Southern California, Los Angeles, CA 90089, USA

² Departments of Computational Medicine and Bioinformatics and Human Genetics,
University of Michigan, Ann Arbor, MI 48109, USA

³ Department of Chemistry and Program in Bioinformatics, Boston University, Boston, MA
02215, USA

⁴ Departments of Chemistry, Physics, and Computer Science, University of Southern
California, Los Angeles, CA 90089, USA

These authors contributed equally to this work.

* To whom correspondence should be addressed. Tel: +1 213 7400552; Fax: +1 213
8214257; Email: rohs@usc.edu

KEYWORDS: Genome structure, DNA shape, nucleosome positioning, transcription
start site, protein-DNA recognition

Supplementary Table 1.

List of current genome assemblies in GBshape.

Common name	Scientific name	Assembly
Mammals		
Human	<i>Homo sapiens</i>	Feb. 2009 (GRCh37/hg19)
Chimp	<i>Pan troglodytes</i>	Feb. 2011 (CSAC 2.1.4/panTro4)
Orangutan	<i>Pongo pygmaeus abelii</i>	July 2007 (WUGSC 2.0.2/ponAbe2)
Gorilla	<i>Gorilla gorilla gorilla</i>	May 2011 (gorGor3.1/gorGor3)
Gibbon	<i>Nomascus leucogenys</i>	Oct. 2012 (GGSC Nleu3.0/nomLeu3)
Rhesus	<i>Macaca mulatta</i>	Oct. 2010 (BGI CR_1.0/rheMac3)
Ferret	<i>Mustela putorius furo</i>	Apr. 2011 (MusPutFur1.0/musFur1)
Bushbaby	<i>Otolemur garnettii</i>	Mar. 2011 (Broad/otoGar3)
Squirrel monkey	<i>Saimiri boliviensis</i>	Oct. 2011 (Broad/saiBol1)
Marmoset	<i>Callithrix jacchus</i>	March 2009 (WUGSC 3.2/calJac3)
Baboon	<i>Papio anubis</i>	Mar. 2012 (Baylor Panu_2.0/papAnu2)
Sloth	<i>Choloepus hoffmanni</i>	Jul. 2008 (Broad/choHof1)
Cat	<i>Felis catus</i>	Sep. 2011 (ICGSC Felis_catus 6.2/felCat5)
Tree shrew	<i>Tupaia belangeri</i>	Dec. 2006 (Broad/tupBel1)
Panda	<i>Ailuropoda melanoleuca</i>	Dec. 2009 (BGI-Shenzhen 1.0/ailMel1)
Dog	<i>Canis lupus familiaris</i>	Sep. 2011 (Broad CanFam3.1/canFam3)
Shrew	<i>Sorex araneus</i>	Aug. 2008 (Broad/sorAra2)
Cow	<i>Bos taurus</i>	Oct. 2011 (Baylor Btau_4.6.1/bosTau7)
Sheep	<i>Ovis aries</i>	Aug. 2012 (ISGC Oar_v3.1/oviAri3)
Mouse lemur	<i>Microcebus murinus</i>	Jul. 2007 (Broad/micMur1)
Kangaroo rat	<i>Dipodomys ordii</i>	Jul. 2008 (Broad/dipOrd1)
Minke whale	<i>Balaenoptera acutorostrata scammoni</i>	Oct. 2013 (BalAcu1.0/balAcu1)
Megabat	<i>Pteropus vampyrus</i>	Jul. 2008 (Broad/pteVam1)
Squirrel	<i>Spermophilus tridecemlineatus</i>	Nov. 2011 (Broad/speTri2)
Rock hyrax	<i>Procavia capensis</i>	Jul. 2008 (Broad/proCap1)
Alpaca	<i>Vicugna pacos</i>	Mar. 2013 (Vicugna_pacos-2.0.1/vicPac2)
Dolphin	<i>Tursiops truncatus</i>	Oct. 2011 (Baylor Ttru_1.4/turTru2)
Microbat	<i>Myotis lucifugus</i>	Jul. 2010 (Broad Institute Myoluc2.0/myoLuc2)
Pig	<i>Sus scrofa</i>	Aug. 2011 (SGSC Sscrofa10.2/susScr3)
Naked mole-rat	<i>Heterocephalus glaber</i>	Jan. 2012 (Broad HetGla_female_1.0/hetGla2)
Guinea pig	<i>Cavia porcellus</i>	Feb. 2008 (Broad/cavPor3)
Tarsier	<i>Tarsius syrichta</i>	Aug. 2008 (Broad/tarSyr1)
Pika	<i>Ochotona princeps</i>	May 2012 (OchPri3.0/ochPri3)
Horse	<i>Equus caballus</i>	Sep. 2007 (Broad/equCab2)
Mouse	<i>Mus musculus</i>	Dec. 2011 (GRCm38/mm10)
Chinese hamster	<i>Cricetulus griseus</i>	Jul. 2013 (C_griseus_v1.0/criGri1)
White rhinoceros	<i>Ceratotherium simum</i>	May 2012 (CerSimSim1.0/cerSim1)
Rat	<i>Rattus norvegicus</i>	Mar. 2012 (RGSC 5.0/rn5)
Armadillo	<i>Dasybus novemcinctus</i>	Dec. 2011 (Baylor/dasNov3)
Elephant	<i>Loxodonta africana</i>	Jul. 2009 (Broad/loxAfr3)
Manatee	<i>Trichechus manatus latirostris</i>	Oct. 2011 (Broad v1.0/triMan1)
Rabbit	<i>Oryctolagus cuniculus</i>	Apr. 2009 (Broad/oryCun2)
Tenrec	<i>Echinops telfairi</i>	Nov. 2012 (Broad/echTel2)
Tasmanian devil	<i>Sarcophilus harrisii</i>	Feb. 2011 (WTSI Devil_ref v7.0/sarHar1)
Hedgehog	<i>Erinaceus europaeus</i>	May 2012 (EriEur2.0/eriEur2)
Platypus	<i>Ornithorhynchus anatinus</i>	Mar. 2007 (WUGSC 5.0.1/ornAna1)
Wallaby	<i>Macropus eugenii</i>	Sep. 2009 (TWGS Meug_1.1/macEug2)
Vertebrates		
Chicken	<i>Gallus gallus</i>	Nov. 2011 (ICGSC Gallus_gallus-4.0/galGal4)
Turkey	<i>Meleagris gallopavo</i>	Dec. 2009 (TGC Turkey_2.01/melGal1)
Budgerigar	<i>Melopsittacus undulatus</i>	Sep. 2011 (WUSTL v6.3/melUnd1)

	Zebra finch	<i>Taeniopygia guttata</i>	Feb. 2013 (WashU taeGut324/taeGut2)
	Medium ground finch	<i>Geospiza fortis</i>	Apr. 2012 (GeoFor_1.0/geoFor1)
	American alligator	<i>Alligator mississippiensis</i>	Aug. 2012 (allMis0.2/allMis1)
	Elephant shark	<i>Callorhynchus milii</i>	Dec. 2013 (Callorhynchus_milii-6.1.3/calMil1)
	Lizard	<i>Anolis carolinensis</i>	May 2010 (Broad AnoCar2.0/anoCar2)
	Painted turtle	<i>Chrysemys picta bellii</i>	Dec. 2011 (v3.0.1/chrPic1)
	<i>X. tropicalis</i>	<i>Xenopus tropicalis</i>	Nov. 2009 (JGI 4.2/xenTro3)
	Coelacanth	<i>Latimeria chalumnae</i>	Aug. 2011 (Broad/latCha1)
	Nile tilapia	<i>Oreochromis niloticus</i>	Jan. 2011 (Broad oreNil1.1/oreNil2)
	Zebrafish	<i>Danio rerio</i>	Jul. 2010 (Zv9/danRer7)
	Tetraodon	<i>Tetraodon nigroviridis</i>	Mar. 2007 (Genoscope 8.0/tetNig2)
	Medaka	<i>Oryzias latipes</i>	Oct. 2005 (NIG/UT MEDAKA1/oryLat2)
	Fugu	<i>Takifugu rubripes</i>	Oct. 2011 (FUGU5/fr3)
	Stickleback	<i>Gasterosteus aculeatus</i>	Feb. 2006 (Broad/gasAcu1)
	Atlantic cod	<i>Gadus morhua</i>	May 2010 (Genofisk GadMor_May2010/gadMor1)
	Lamprey	<i>Petromyzon marinus</i>	Sep. 2010 (WUGSC 7.0/petMar2)
Deuterostomes	Lancelet	<i>Branchiostoma floridae</i>	Mar. 2006 (JGI 1.0/braFlo1)
	<i>C. intestinalis</i>	<i>Ciona intestinalis</i>	Mar. 2005 (JGI 2.1/ci2)
	<i>S. purpuratus</i>	<i>Strongylocentrotus purpuratus</i>	Sep. 2006 (Baylor 2.1/strPur2)
Insects	<i>D. melanogaster</i>	<i>Drosophila melanogaster</i>	Apr. 2006 (BDGP R5/dm3)
	<i>D. simulans</i>	<i>Drosophila simulans</i>	Apr. 2005 (WUGSC mosaic 1.0/droSim1)
	<i>D. sechellia</i>	<i>Drosophila sechellia</i>	Oct. 2005 (Broad/droSec1)
	<i>D. erecta</i>	<i>Drosophila erecta</i>	Aug. 2005 (Agencourt prelim/droEre1)
	<i>D. yakuba</i>	<i>Drosophila yakuba</i>	Nov. 2005 (WUGSC 7.1/droYak2)
	<i>D. ananassae</i>	<i>Drosophila ananassae</i>	Aug. 2005 (Agencourt prelim/droAna2)
	<i>D. pseudoobscura</i>	<i>Drosophila pseudoobscura</i>	Nov. 2004 (FlyBase 1.03/dp3)
	<i>D. pseudoobscura</i>	<i>Drosophila pseudoobscura</i>	Feb. 2006 (Baylor CAF1/dp4)
	<i>D. persimilis</i>	<i>Drosophila persimilis</i>	Oct. 2005 (Broad/droPer1)
	<i>D. virilis</i>	<i>Drosophila virilis</i>	Aug. 2005 (Agencourt prelim/droVir2)
	<i>D. mojavensis</i>	<i>Drosophila mojavensis</i>	Aug. 2005 (Agencourt prelim/droMoj2)
	<i>D. grimshawi</i>	<i>Drosophila grimshawi</i>	Aug. 2005 (Agencourt prelim/droGri1)
	<i>A. mellifera</i>	<i>Apis mellifera</i>	Jan. 2005 (Baylor 2.0/apiMel2)
	<i>A. gambiae</i>	<i>Anopheles gambiae</i>	Feb. 2003 (IAGEC MOZ2/anoGam1)
Nematodes	<i>C. elegans</i>	<i>Caenorhabditis elegans</i>	Oct. 2010 (WS220/ce10)
	<i>C. briggsae</i>	<i>Caenorhabditis briggsae</i>	Jan. 2007 (WUGSC 1.0/cb3)
	<i>C. remanei</i>	<i>Caenorhabditis remanei</i>	May 2007 (WUGSC 15.0.1/caeRem3)
	<i>C. brenneri</i>	<i>Caenorhabditis brenneri</i>	Feb. 2008 (WUGSC 6.0.1/caePb2)
	<i>C. japonica</i>	<i>Caenorhabditis japonica</i>	Mar. 2008 (WUGSC 3.0.2/caeJap1)
	<i>P. pacificus</i>	<i>Pristionchus pacificus</i>	Feb. 2007 (WUGSC 5.0/priPac1)
Fungi	<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>	Apr. 2011 (SacCer_Apr2011/sacCer3)
Plants	<i>A. thaliana</i>	<i>Arabidopsis thaliana</i>	May. 2011 (TAIR10/araTha10)
Protists	<i>P. falciparum</i>	<i>Plasmodium falciparum</i>	Jun. 2007 (PlaFal_5.5/plaFal1)
Bacteria	<i>E. coli K12</i>	<i>Escherichia coli K12</i>	2005 (MG1655/eschColi_K12)
Others	Sea hare	<i>Aplysia californica</i>	Sept. 2008 (Broad 2.0/aplCal1)

Supplementary Table 2.

List of current assemblies supported by BLAT sequence search in GBshape.

Common name	Scientific name	Assembly
Mammals		
Mouse	<i>Mus musculus</i>	Dec. 2011 (GRCm38/mm10)
Rat	<i>Rattus norvegicus</i>	Mar. 2012 (RGSC 5.0/rn5)
Insects		
<i>D. melanogaster</i>	<i>Drosophila melanogaster</i>	Apr. 2006 (BDGP R5/dm3)
<i>D. simulans</i>	<i>Drosophila simulans</i>	Apr. 2005 (WUGSC mosaic 1.0/droSim1)
<i>D. sechellia</i>	<i>Drosophila sechellia</i>	Oct. 2005 (Broad/droSec1)
<i>D. pseudoobscura</i>	<i>Drosophila pseudoobscura</i>	Feb. 2006 (Baylor CAF1/dp4)
Nematodes		
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>	Oct. 2010 (WS220/ce10)
Fungi		
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>	Apr. 2011 (SacCer_Apr2011/sacCer3)

Author contributions

T.-P.C. developed the methodology for generating whole-genome annotation tracks for diverse DNA shape features, designed and developed the GBshape database and its extendable high-throughput prediction platform, and analyzed transcription start sites. L.Y. analyzed nucleosome-binding sites and provided advice in the initial stage of the database development. T.Z. developed the high-throughput method for DNA shape predictions. S.C.J.P. and T.D.T. developed the methodology for hydroxyl radical cleavage annotations, developed the ORChID database, and provided the modENCODE data for nucleosome-bound sequences. B.J.M. and S.V.N. derived the methodology for TSS calling and provided the *Drosophila* TSS sequence data. T.-P.C. and R.R. wrote the manuscript with input from all authors. R.R. conceived, designed, and supervised the project.