

iDog

Michael Armas and Jonar Cowan
Department of Biology
BIOL 375

What is iDog?

- Integrated research resource for domestic dogs and wild canids
- Based out of China
- Several data types:
 - Genes
 - Genomes
 - SNPs
 - Breed/Disease Traits
 - Gene Expression



What is iDog?

- Links to other specific databases
 - Connected to NGDC (which links to over 40 databases)
- Public release on January 17th, 2017
- Updated every few months
 - Last updated June 21st, 2019

Database Specifics

- Phenotypic data gathered from Kennel Clubs/Unions
- Curated from dog research communities and public resources
 - CIDD
 - OIMA
 - TheDogPlace
 - Ensembl
 - UniProt
- Maintained by National Genomics Data Center
- Funded by various associations and programs of the Chinese Academy of Science

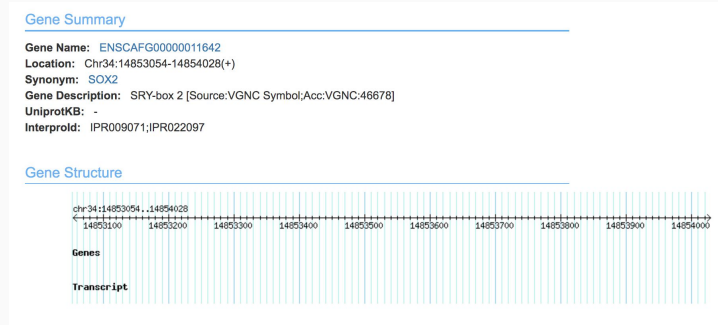
Contents

- Over 60 million data points
- Provides literature about specific species of canids
- Provides tools for genomic data visualization and analysis

Data contents	Data statistics
Gene	
Gene integrated from Ensembl database	32 220
Gene associated with Dog disease	229
SNP	
Individuals	127
Non-redundant SNPs identified	42 871 184
Non-redundant SNPs annotated in gene	22 031 720
Phenotype & disease	
Standard Breeds	473
Diseases	783
Genotype-phenotype pairs (G2Ps)	594
Genome	
Scaffolds for wolf	581
Scaffolds for Dhole	749
Gene expression	
RNA-Seq projects	7
Experiments	83
Tissues	5
Genes well annotated	27 534
Gene ontology (GO)	
Molecular function	
Genes	14 361
Annotations	60 030
Biological process	
Genes	15 079
Annotations	103 120
Cellular component	
Genes	15 884
Annotations	65 214
Literature	
Papers and books	6 535

Utility for the Public

- Targeted Audience has background in biology and genetics
- Some information is understandable by those without experience
- Phenotype database easily understandable



Breeds Search ▾

Breed Name ▾

Temperament ▾

Affectionate Friendly Alert
 Intelligent Loyal Calm
 Lively Smart Courageous
 Gentle Confident
 Independent Devoted
 Fearless Playful Dignified
 Aggressive Outgoing

Size ▾

All Small Medium
 Large

Trainability ▾

All Responds Well
 Eager to Please Independent
 Agreeable May be Stubborn
 Easy Training Controlled

Weight ▾

1 ~ 105 kg

Height ▾

12 ~ 90 cm

Life Expectancy ▾

5 ~ 20 years

User Friendliness

- The layout design is organized, but lacks some genetic information compared to other databases.
- The database definitions can be found when searching up genes to provide basic information
- Navigating this site could be difficult if user does not understand biological terminology
- Tutorials are not helpful

The screenshot displays the iDog website interface. At the top, there is a navigation bar with links for 'Databases', 'Tools', 'Standards', 'Publications', and 'About', along with a 'Sign in' button. Below this is a secondary navigation bar with 'Home', 'Browse', 'Search', 'Tools', 'Download', and 'Documents'. The main content area features a 'Welcome to iDog!' message and a search bar with a dropdown menu for 'Gene Symbol' and a search input field containing 'e.g. ENPP1'. A small image of a dog is positioned to the right of the search bar. Below the search bar, there are several database links represented by icons and text: 'Gene' (Gene and Annotation), 'DogGD' (Dog Genome Database), 'DogSD' (Dog SNP Database), 'DogPD' (Dog Phenotype Database), 'DogED' (Dog Expression Database), 'GO' (GO Function Annotations), 'Dog Online' (Online tools), 'DogHDC' (Dog-Human Disease Connection), and 'Dog Literature' (Dog Literature Query). A 'Publication' section lists a paper by Tang B., Zhou Q., Dong L., et al. (2018). A 'Related Links' section mentions the 'Dog 10K genomes project (dog10K)'. On the right side, there are two panels: 'Data Statistics' showing various counts (e.g., 32,220 Genes, 229 Dog disease genes) and 'What's New?' listing recent updates (e.g., CTVT Genotype-monomorphic Sites File is now available to access).

User Friendliness (cont.)


- The search function only works to find gene symbol, gene name, and gene location
- Example searches are provided for quick understanding
- However, a general search bar is not found and requires you to look for tutorials/help

Welcome to iDog !

iDog, an integrated resource for domestic dog (*Canis lupus familiaris*) and wild canids, provides the worldwide dog research community a variety of data services. This includes Genes, Genomes, SNPs, Breed/Disease Traits, Gene Expressions, GO Function Annotations, Dog-Human Homolog Diseases and Literatures. In addition, iDog provides Online tools for performing genomic data visualization and analyses.

Search:

e.g. MERTK, ENSCAFG00000000338, Chr1:14702150-37601000(+)



FAQ/Tutorials

- Finding the help section is difficult and offers a limited amount of information
- Tutorials are too little info, FAQ is too much

Tutorial

1. Gene Query

Users can use *filter* restricting their gene queries with our gene query form. Detailed information of the results in iDog are presented in 10 categories, including Gene Summary, Gene Structure, SNP Information, Exon Information, Exon Genome Assembly, GMA3 Catalog, Dog Disease Information, Expression Information, Homology Gene, Gene Ontology and Genome Browser.

Click here for a detailed tutorial.

2. Dog Genome Database

Dog Genome Database (DogGD) is a data container for the genome assembly information of worldwide genomes. It features scaffold, annotated gene and protein information.

Detailed tutorials are provided below.

- Browse data
 - Scaffold
 - Gene
- Search data
 - Simple Search
 - Blast
- Genome View
 - Download

3. Dog SNP Database

Dog Genome SNP Database (DogSD) is a data container for the variation information of dog/wolf genomes. In the current version, DogSD contains 42 million non-redundant SNPs which curated from 127 individual samples and doSNP² (ver 146).

Detailed tutorials are provided below:

- Browse SNP data
 - SNP tables
 - SNPs in Gene
 - SNPs on Chromosome
- Search data
 - Single individual search
 - Compared individuals search
- Download

5. Dog Expression Database

Dog Expression Database (DogED) is a repository of gene expression profiles derived entirely from RNA-Seq data analysis of tissues from Canis. Currently, 7 RNA-Seq Projects and 83 Experiments are collected from NCBI.

Detailed tutorials are provide below.

- Browse Data
 - RNA-Seq Project
 - Gene Differential Expression in Project
 - Annotated Gene
- Search Data
 - Gene Expression Search

6. Dog Human Disease Connection

Dog Human Disease Connection (DogHDC) is dedicated to provide homology information for dog disease associated genes in other species, especially in human. All homology information is curated from NCBI Homology Gene as well as OMIM.

Detailed tutorials are provided below:

- Browse Data
 - Dog & Human homology information
- Search Data
 - Dog & Human homology information

7. Dog Online

Dog online provides data visualization and analysis tools for iDog. Users can analyze their own data and visualize the results online.

Detailed tutorials are provided below.

- Data Visualization
 - Dog SNP Genome Browser
- Data Analysis
 - BWA

FAQ

How to find interesting gene? FAQ

[Dog Genome Database \(DogGD\)](#)

1. How to search annotated information in DogGD? FAQ
2. How perform blast with my own data against DogGD? FAQ

[Dog SNP Database \(DogSD\)](#)

1. How to retrieve RefSNP information? FAQ
2. How to retrieve Individual SNP information? FAQ
3. How to find the SNP annotated information? FAQ
4. How to search for SNP in single individual? FAQ
5. How to search for SNP in multiple individuals? FAQ
6. How to download data of DogSD? FAQ

[Dog Phenotype Database \(DogPD\)](#)

1. How to search breed information? FAQ
2. How to search dog disease information ? FAQ

[Dog Expression Database \(DogED\)](#)

1. How to find gene profiles by FPKM value? FAQ

[Dog Human Disease Connection \(DogHDC\)](#)

1. How to find homology information of dog and human? FAQ

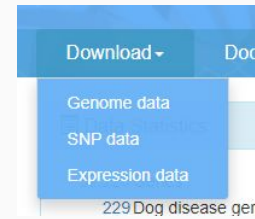
Downloadable Data

- The downloadable files hold a lot of memory, which results in a slow download time
- Specific Programs are needed to run downloaded files (.fa and .bam extensions types)
- Converting these files is problematic

All Gene Expression FTP

Gene expression in each project

Name	Gene expression profile in project	Differential gene profile within project
PRJNA271278	FTP	FTP
PRJNA276284	FTP	FTP
PRJNA297808	FTP	FTP
PRJNA314784	FTP	FTP
PRJNA360981	FTP	FTP
PRJNA382537	FTP	FTP
PRJNA393099	FTP	FTP



Download Sequence Data		
File Name	MD5	Download
Druid Genome File	8789956a2264e30a07f93aa11c2e	FTP
Druid CDS File	4f71893894949f1f4a2c01c81a2a	FTP
Druid Protein File	48f2561713d6d8f8156151514	FTP
Woff Genome File	78333d5e11996130a88888888	FTP
Woff CDS File	0f18411014765a36d16a18a85a	FTP
Woff Protein File	63a2666a7936c1985651c4e41193	FTP
CTVT Genome-monomorphic Sites File	9a8007181a14d2761378107071a	FTP

Download Annotation Data		
File Name	MD5	Download
Druid Annotation File	0c8834e57343201f1c0381959a0	FTP
Woff Annotation File	7461a8b3a6027743ca1a6d071a1f	FTP

Data list

1. bam file
2. bai file
3. fastq file
4. vcf file

Download data

1.bam file

Note : The following list provides the links to several individual samples. Although the links can be used to download the data directly, due to the size of the datasets, we recommend downloading the data using the following instructions:

Under linux environment: use wget command
 such as `wget ftp://download.big.ac.cn/100g/00g5/bam/1735.rmdup.realign.recalibration.bam`

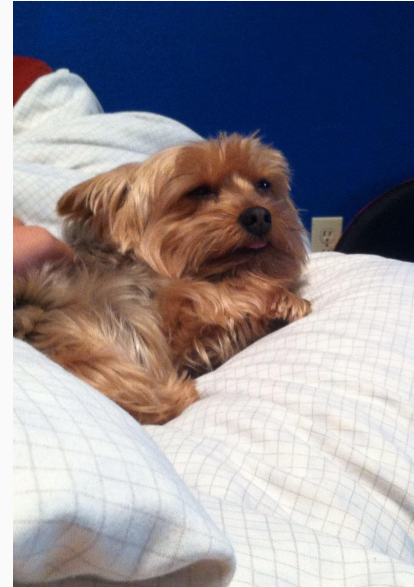
Under windows environment: use FTP tool(ftp://)

host: `download.big.ac.cn` username: anonymous password: anonymous
 If you connect successfully, then change to the path dogs.

File name	MD5
1735.rmdup.realign.recalibration.bam	b0e93009103f0e67989c76934d3db59e
2972.rmdup.realign.recalibration.bam	15e3e870f08f77923edfb00a29e6d2a
4669.rmdup.realign.recalibration.bam	eff542b163eac03582d77d0a9f009936
BA19.rmdup.realign.recalibration.bam	e74b1b44ef7e387b2852e13bb0200bec2
Basenji.rmdup.realign.recalibration.bam	124a395a7196fb2cb8fe30a050c3a0d
Chihuahua.rmdup.realign.recalibration.bam	446184b40b8b4d89c4a0c4c6

Demo

- Sparky is a 13 yr. Old Yorkshire Terrier
- Lately has developed cloudiness in his eyes



Judgement

- This database is not useful or easy to navigate if a person is not familiar with the field of study
- Data is mainly about dogs, but presented in a professional fashion
 - Curated from many credible databases
 - Published in credible journal
 - Mainly useful for scientists, geneticists, dog owners

