

Resource Summary Report

Generated by SPARC SAWG on Aug 9, 2026

Non-Human Genome Segmental Duplication Database

RRID:SCR_000470

Type: Tool

Proper Citation

Non-Human Genome Segmental Duplication Database (RRID:SCR_000470)

Resource Information

URL: <http://projects.tcag.ca/xenodup/>

Proper Citation: Non-Human Genome Segmental Duplication Database (RRID:SCR_000470)

Description: THIS RESOURCE IS NO LONGER IN SERVICE, documented on July 16, 2013. It contains information about segmental duplications in the genomes of chimpanzee, mouse, and rat. The criteria used to identify regions of segmental duplication are: * Sequence identity of at least 90% * Sequence length of at least 5 kb * Not be entirely composed of repetitive elements. **BACKGROUND:** The high quality of the mouse genome draft sequence and its associated annotations are an invaluable biological resource. Identifying recent duplications in the mouse genome, especially in regions containing genes, may highlight important events in recent murine evolution. In addition, detecting recent sequence duplications can reveal potentially problematic regions of the genome assembly. We use BLAST-based computational heuristics to identify large (≥ 5 kb) and recent ($\geq 90\%$ sequence identity) segmental duplications in the mouse genome sequence. Here we present a database of recently duplicated regions of the mouse genome found in the mouse genome sequencing consortium (MGSC) February 2002 and February 2003 assemblies. **RESULTS:** We determined that 33.6 Mb of 2,695 Mb (1.2%) of sequence from the February 2003 mouse genome sequence assembly is involved in recent segmental duplications, which is less than that observed in the human genome (around 3.5-5%). From this dataset, 8.9 Mb (26%) of the duplication content consisted of "unmapped" chromosome sequence. Moreover, we suspect that an additional 18.5 Mb of sequence is involved in duplication artifacts arising from sequence misassignment errors in this genome assembly. By searching for genes that are located within these regions, we identified 675 genes that mapped to duplicated regions of the mouse genome. Sixteen of these genes appear to have been duplicated independently in the human genome. From our dataset we further characterized a 42 kb recent segmental duplication of *Mater*, a maternal-effect gene essential for embryogenesis in mice. **CONCLUSION:** Our results provide an initial analysis of the recently duplicated sequence and gene content of the

mouse genome. Many of these duplicated loci, as well as regions identified to be involved in potential sequence misassignment errors, will require further mapping and sequencing to achieve accuracy. A Genome Browser database was set up to display the identified duplication content presented in this work. This data will also be relevant to the growing number of investigators who use the draft genome sequence for experimental design and analysis. The segmental duplication data and summary statistics are available for download and can also be visualized in a genome browser in the GBrowse section. Selected annotation tracks (except the segmental duplication track) have also been obtained from UCSC and loaded into the genome browser. Detailed information (e.g. overlapping genes, overlapping clones, detailed alignment) can be obtained by clicking on a duplication cluster in GBrowse. Both keyword search and BLAT search are available. Analyses based on previous genome assemblies can be found in the Previous Analyses section. Recent Developments The Non-Human Genome Segmental Duplication Database is continually updated including the archived copies of the analysis of all previous genome assemblies and will include all new species as they become available. Acknowledgments We thank The Centre for Applied Genomics at the Hospital for Sick Children (HSC) as well as collaborators worldwide. Supported by Genome Canada the Howard Hughes Medical Institute International Scholar Program (to S.W.S.) and the HSC Foundation.

Synonyms: Non-Human Genome Segmental Duplication Database

Resource Type: database, data or information resource

Keywords: genes, genome, chicken, chimpanzee, chromosome, dna, dog, mouse, rat, segmental duplications

Availability: THIS RESOURCE IS NO LONGER IN SERVICE

Resource Name: Non-Human Genome Segmental Duplication Database

Resource ID: SCR_000470

Alternate IDs: nif-0000-03194

Record Creation Time: 20220129T080201+0000

Record Last Update: 20260808T190342+0000

Ratings and Alerts

No rating or validation information has been found for Non-Human Genome Segmental Duplication Database.

No alerts have been found for Non-Human Genome Segmental Duplication Database.

Data and Source Information

Source: [SciCrunch Registry](#)

Usage and Citation Metrics

We found 1 mentions in open access literature.

Listed below are recent publications. The full list is available at [SPARC SAWG](#) .

Feuk L, et al. (2005) Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. PLoS genetics, 1(4), e56.