

DATABASE

Open Access



TeCD: The eccDNA Collection Database for extrachromosomal circular DNA

Jing Guo^{1,2,3,4†}, Ze Zhang^{1,2,5†}, Qingcui Li⁵, Xiao Chang^{3*} and Xiaoping Liu^{1,2,4*}

Abstract

Background Extrachromosomal circular DNA (eccDNA) is a kind of DNA that widely exists in eukaryotic cells. Studies in recent years have shown that eccDNA is often enriched during tumors and aging, and participates in the development of cell physiological activities in a special way, so people have paid more and more attention to the eccDNA, and it has also become a critical new topic in modern biological research.

Description We built a database to collect eccDNA, including animals, plants and fungi, and provide researchers with an eccDNA retrieval platform. The collected eccDNAs were processed in a uniform format and classified according to the species to which it belongs and the chromosome of the source. Each eccDNA record contained sequence length, start and end sites on the corresponding chromosome, order of the bases, genomic elements such as genes and transposons, and other information in the respective sequencing experiment. All the data were stored into the TeCD (The eccDNA Collection Database) and the BLAST (Basic Local Alignment Search Tool) sequence alignment function was also added into the database for analyzing the potential eccDNA sequences.

Conclusion We built TeCD, a platform for users to search and obtain eccDNA data, and analyzed the possible potential functions of eccDNA. These findings may provide a basis and direction for researchers to further explore the biological significance of eccDNA in the future.

Keywords Circular DNA, Extrachromosomal DNA, eccDNA

[†]Jing Guo and Ze Zhang contributed equally to this study as the first authors.

*Correspondence:

Xiao Chang
chxlaugh@163.com
Xiaoping Liu
xpliu@ucas.ac.cn

¹ Key Laboratory of Systems Biology, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310013, China

² Key Laboratory of Systems Health Science of Zhejiang Province, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310013, China

³ Institute of Statistics and Applied Mathematics, Anhui University of Finance & Economics, Bengbu 233030, China

⁴ School of Mathematics and Statistics, Shandong University, Weihai 264209, Shandong, China

⁵ School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310013, China

Background

In eukaryotes, a particular class of circular DNA molecules is separated from the normal genome, dissociated from the chromosome, and participates in physiological or pathological processes with a special way [1]. Because these molecules are independent DNA molecules outside the chromosome, and, they are often circular, so they are called extrachromosomal circular DNA, or eccDNA [2].

A lot of studies have shown that eccDNA can be involved in a wide range of biological processes [3–5]. For example, in cancer, eccDNA is the main factor leading to tumor heterogeneity and is regarded as a marker of genomic instability. However, eccDNA is also detected in healthy cells and can appear in a tissue-specific or developmentally regulated mode, which implies possible physiological effects [3]. Although researchers have found a large amount of eccDNA in normal and cancer cells



[6, 7], people are still paying more and more attention to the connection between eccDNA and cancer [8–10]. Many studies have shown that eccDNA plays an important role in genetic variation, evolution, genomic instability, genomic plasticity, drug resistance, environmental adaptation, mutation and tumorigenesis [8]. An eccDNA usually contains tumor genes or drug resistance genes [9–13], so it can promote cancer cells to gain growth or survival advantages, and it can also lead to gene deletion, mutation, duplication, amplification or migration, intercellular genetic heterogeneity, and adaptive evolution [9]. Through the study of colon cancer HT29 cells, ovarian cancer, and breast cancer tissues, researchers found that eccDNA may promote the drug resistance of cancer or tumor cells by amplifying tumor genes [10]. One characteristic of the plasticity of the genome is the presence of eccDNA [12], which is the ability of the genome to produce different phenotypes based on different environmental cues and is related to a variety of human diseases and related phenotypes. High levels of eccDNA are related to genome instability, exposure to carcinogens, and cell aging [11]. The study of Fanconi's anemia in DNA repair deficiency syndrome has shown genetic instability and increased levels of associated eccDNA molecules [13]. In summary, researchers are still exploring and studying the functions of eccDNA, and there is still a lot of unknown information that they have not grasped. The research of eccDNA may have a certain impact on traditional genetics and genomics in the future.

However, the integrative resources that already exist for eccDNA, such as eccDNAdb [14] and CircleBase [15], contain only human data or specifically human tumor data as far as we know. We would like to build a database platform containing eccDNA from a variety of eukaryotes. Thus, we established a database named TeCD (The eccDNA Collection Database) to integrate the eccDNA data scattered in public literature. The species distribution of the data in TeCD covers animals, plants and fungi. Researchers can obtain the genomic elements or locations of eccDNA and detect the related sequences by BLAST (Basic Local Alignment Search Tool) from TeCD.

Construction and content

Database implementation

The web server was deployed based on Nginx 1.18.0 (<http://nginx.org/>) and uWSGI 2.0.19.1 (<https://pypi.org/project/uWSGI/>), and the Django 2.2.12 framework (<https://www.djangoproject.com/>) was used to implement Python language interface on a back-end server. And all data were stored in MySQL 8.0.26 database (<https://dev.mysql.com/>) for Linux on x86_64 (Ubuntu Server 20.04.2 LTS). The web page templates were used on Semantic UI framework (<https://semantic-ui.com/>),

DataTables (<http://datatables.net>), and jQuery (<http://jquery.com>) to establish a user-friendly front-end interface. HTML, CSS, and JavaScript were selected as the client-side languages for front-end design (Fig. 1).

Data collection

Many studies have proven that eccDNA comes from the chromosome, not from any pre-existing extrachromosomal precursors [16, 17], so each detected eccDNA can be matched to a corresponding site on chromosome. At present, most the studies about eccDNA only focus on both the main characteristics of eccDNA (such as length, structure, and expressive characteristics) and the development of isolating methods that can be used to detect eccDNA [18, 19]. The research for the function of eccDNA is still few. To facilitate the further exploration of the function of eccDNA, we collected the existing information of eccDNA and built a database to organize them for acquirement.

The eccDNA data were collected from the published literature related to eccDNA for eukaryotes [7, 17, 20–35], and included the information such as eccDNA type, length, locus shedding from chromosomes, sequencing approach and so on. These data are relatively scattered in different eccDNA studies and need further process after the collection of the data (Fig. 1).

Data processing

There were about 2,343,000 eccDNAs to be collected from public literature. The eccDNAs were filtered according to whether they contained the start and end sites of eccDNA on the corresponding chromosome, and a total of 1,846,905 eccDNAs were obtained. We matched the eccDNAs through their start and end sites to the reference genome (Homo sapiens GRCh38.p13, Saccharomyces cerevisiae S288C, Arabidopsis thaliana TAIR10.1, Gallus gallus bGalGal1.mat.broiler. GRCg7b and Mus musculus GRCm39) from the National Center for Biotechnology Information (NCBI, <https://www.ncbi.nlm.nih.gov/>). Then, the filtered data were divided into groups according to species (Table 1). Then the repeated sequences were identified by BLAST tool [36] with a threshold of e-value 10^{-6} in individual species, and were removed from the database. After the filter, we obtained a total of 948,469 eccDNAs. The amount of eccDNA for each species is shown in Table 1.

In this database, we assigned a unique identification ID which is composed of species information and alphanumeric ordering to each eccDNA. For example, ecc_Hs_100000001 is the ID of eccDNA from Homo sapiens (where 'Hs' represents Homo sapiens), and ecc_Sc_100000001 is the ID of eccDNA from Saccharomyces cerevisiae (where 'Sc' represents Saccharomyces

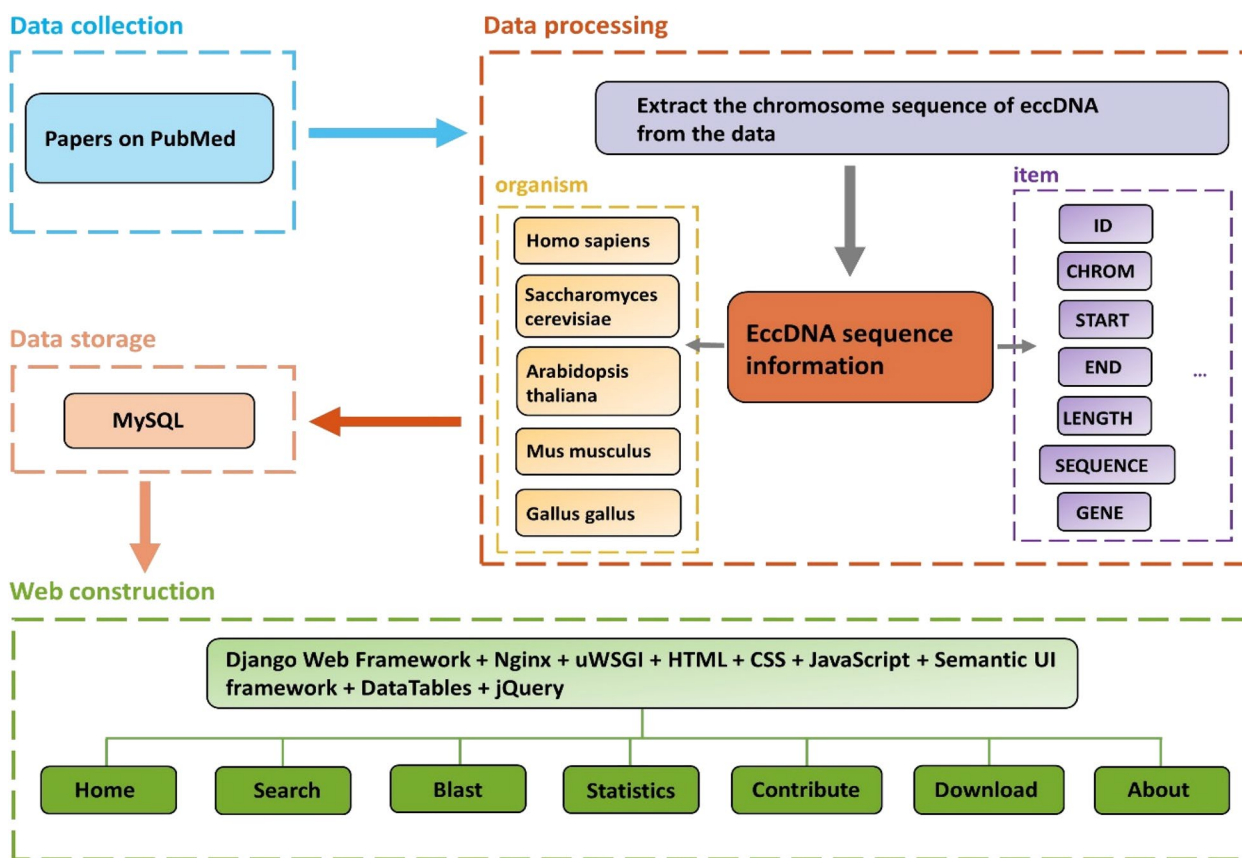


Fig. 1 The flow chart of TeCD. The raw data were collected from published literature. The chromosome sequence corresponding to eccDNA was obtained according to the locus. The gene, locus information, number of reads variants, and other information were extracted from the data and stored in MySQL 8.0.26 database. The information could be browsed and retrieved according to organism and item. Finally, the TeCD website was built with Django web framework for users

Table 1 The amount of eccDNA for each organism

	Homo sapiens	Saccharomyces cerevisiae	Arabidopsis thaliana	Gallus gallus	Mus musculus
Obtained data	379,148	2111	702	1,087,503	377,441
After BLAST	193,294	1266	591	433,993	319,325

cerevisiae). Besides, we also marked the source of each eccDNA based on the tissue and cell type of the sample, and statistically showed that there is a total of 42 categories for the source of tissue and cell type in the database. In addition, each eccDNA was marked the potential source of chromosomes and start/end sites in the chromosome. These classifications can help any scientific researcher to accurately retrieve information on the TeCD platform and carry out research work.

Next, each eccDNA was accurately matched to the corresponding genes or partially covered genes by searching the NCBI genome database. In order

to facilitate scientific researchers for obtaining relevant gene information instantly when browsing and retrieving eccDNA, we have made a direct link to each matched gene, and NCBI’s gene bank will provide a comprehensive and authoritative explanation for the gene (Fig. 1). Specially, the Arabidopsis gene models are linked to TAIR (The Arabidopsis Information Resource <https://www.arabidopsis.org/>), the most authoritative Arabidopsis genome database and Arabidopsis genome annotation system in the world [37, 38]. Besides, the transposable elements information of the sequence is also included in TeCD.

Classification and statistics of eccDNA

The tissue and cell type to which the sample belongs is an important indicator of eccDNA classification, according to which 948,469 eccDNA records in TeCD could be summarized into 42 types (Fig. 2a). And Fig. 2b-e show sample types for five organisms. The inner circle color of each figure corresponds to the inner circle color of Fig. 2a. Samples from *Saccharomyces cerevisiae* are from 11 categories which consisted of the CEN.PK strain and the S288C strain, and we named samples which from CEN.PK strains as ‘CEN.PK’ and further classified samples which from the S288C strain [21, 23]. In the S288C *Saccharomyces cerevisiae* (Fig. 2b), ‘*MATa Gal2* isogenic WT population’ and ‘*MATa ura3* isogenic WT population’ are two wild-type *Saccharomyces cerevisiae* respectively and are in contrast to ‘*MATa his3Δ1* gene deletion yeast’, ‘*MATa leu2Δ0* gene deletion yeast’, ‘*MATa met15Δ0* gene deletion yeast’ and ‘*MATa ura3Δ0* gene deletion yeast’. ‘*MATa his3Δ1* gene deletion yeast’, ‘*MATa leu2Δ0* gene deletion yeast’, ‘*MATa met15Δ0* gene deletion yeast’ and ‘*MATa ura3Δ0* gene deletion yeast’ are single-gene deletion mutants without *his3Δ1*, *leu2Δ0*, *met15Δ0* and *ura3Δ0*. ‘*MATa his3Δ1* gene deletion with Zeocin’, ‘*MATa*

leu2Δ0 gene deletion with Zeocin’, ‘*MATa met15Δ0* gene deletion with Zeocin’ and ‘*MATa ura3Δ0* gene deletion with Zeocin’ are the samples in which DNA-damaging agent Zeocin is added to the corresponding single-gene deletion mutants [23]. *MATa* is a mating type of *Saccharomyces cerevisiae* [39]. The samples from *Arabidopsis thaliana* contain 4 categories [34], which are leaf, stem, flower, root (Fig. 2c). Notably, each eccDNA of *Arabidopsis thaliana* was detected from one or more samples. We intersected the number of eccDNA detected in different sample types and performed a Wayne diagram analysis of the results (Fig. 2g). The different colored circles represent the samples of different parts of *Arabidopsis*. As shown in Fig.2 h, 381 eccDNAs were detected in 2 samples, 25 eccDNAs were detected in 3 samples, and 5 eccDNAs were detected in 4 samples. The samples from *Gallus gallus* contain 9 categories (Fig. 2d). These samples were derived from mutants of the DT-40 cell line [35]. For example, ‘BRCA1’ means DT40-Chicken cell line-mutated BRCA1. Samples from *Homo sapiens* come from 8 categories (Fig. 2e), of which ‘muscle’ represents human samples from healthy human muscle tissue, ‘leukocytes’ means leukocytes collected after centrifugation

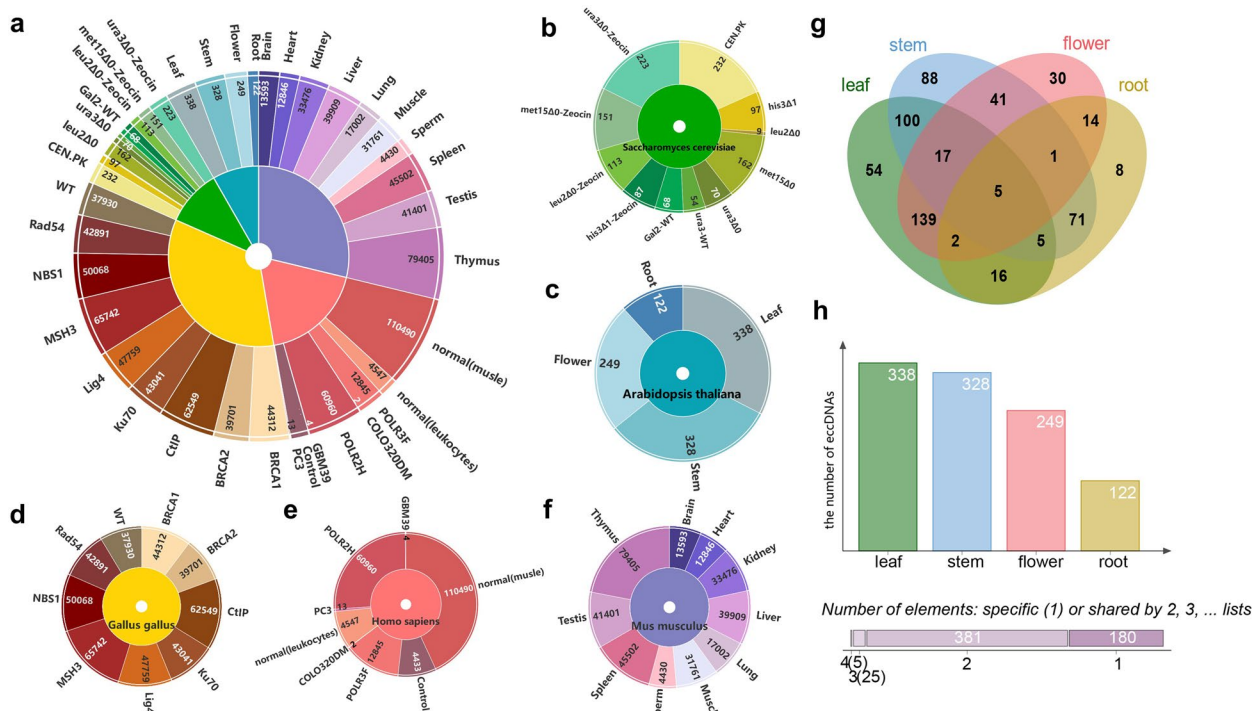


Fig. 2 Sample types for different tissues and cells. **a** Sample types for all organisms and the corresponding number of eccDNA. The 5 colors of the inner ring correspond to 5 organisms. The outer ring represents the sample type for each organism. **b-f** Sample types of *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Gallus gallus*, *Homo sapiens*, and *Mus musculus*. And the color of its inner ring corresponds to the inner ring of the (a). **g** The number of eccDNA shared by samples from different parts of *Arabidopsis*. The different colored circles represent samples of different parts of *Arabidopsis*. **h** The figure above shows the number of eccDNA from four samples of *Arabidopsis*. The figure below shows the number of eccDNA that belongs to 1–4 different samples in common

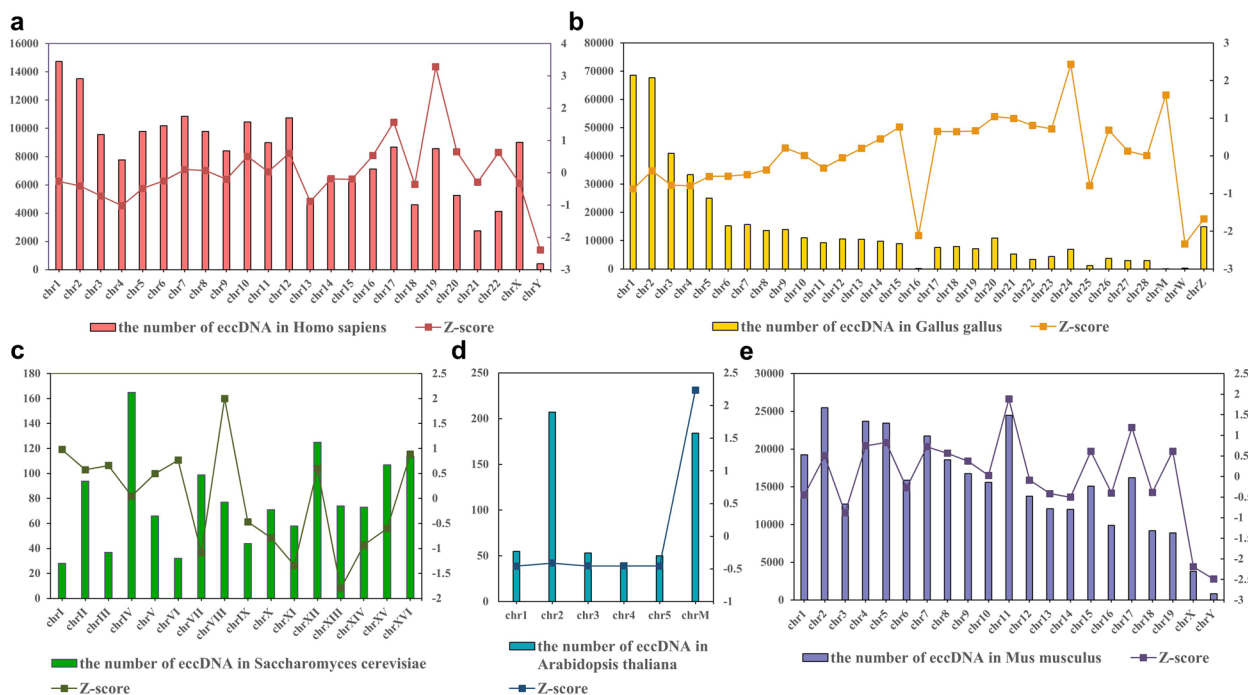


Fig. 3 Data distribution plot. **a-e** The histogram shows the quantitative distribution of eccDNA of different organisms on chromosomes. These species are Homo sapiens, Gallus gallus, Saccharomyces cerevisiae, Arabidopsis thaliana and Mus musculus in turn. The line chart shows the ratio standardized with Z-score. And this ratio is the number of eccDNA in the chromosome divided by the length of the chromosome

from the venous blood (40 ml) of healthy human arm veins [20]. The human embryonic kidney (HEK293) cells are easily transfected and ‘HEK293’ is the control of ‘HEK293 (+POLR2H)’ and ‘HEK293 (+POLR3F)’. ‘HEK293 (+POLR2H)’ means that the POLR2H (RNA Polymerase II, I And III Subunit H) protein-coding gene was transfected into HEK293 cells and ‘HEK293 (+POLR3F)’ means that the POLR3F (RNA Polymerase III Subunit F) protein-coding gene was transfected into HEK293 cells [22]. There is also human prostate cancer cell line ‘PC3’, colon cancer cell line ‘COLO320DM’, and human glioblastoma ‘GBM39’ tumor spheroids derived from patient tissues [24]. The samples from Mus musculus contain 10 categories (Fig. 2d). All samples are from organs or tissues of 6-months-old adult mice, including brain, heart, kidney, liver, lung, muscle, sperm, spleen, testis and thymus [35].

Additionally, we recorded the quantitative distribution of eccDNA on chromosomes, where the histogram represents the specific amount of eccDNA on different chromosomes and the line chart represents the standardized data (Fig. 3). As each chromosome has a different length, in order to eliminate the comparative baseline heterogeneity caused by chromosome heterogeneity, we calculated the ratio of the number of eccDNA in chromosomes to the length of chromosomes and standardized

the ratio by Z-score to compare the amount of eccDNA formed by shedding from different chromosomes. The standardized results were plotted into a line chart. The source of eccDNA is dispersed to 22 autosomes, 2 sexual chromosomes, and mitochondria in humans, and the amount of eccDNA from chromosomes 1, 2, 7, 12 are obviously more than other chromosomes. The eccDNA from X chromosome significantly exceeds the eccDNA from Y chromosome (Fig. 3a). From the line chart of human data, it is not difficult to find that the amount of eccDNA on chromosomes 17 and 19 accounts for the largest proportion (Fig. 3a). The genome of Gallus gallus has 39 autosomes and two sex chromosomes, but in the collected data, eccDNA only comes from 28 autosomes and 2 sex chromosomes (Fig. 3b). Mostly, the number of eccDNA on each chromosome of Saccharomyces cerevisiae was 50–150. It was found on the eighth chromosome “ChrVIII” of Saccharomyces cerevisiae, the ratio of eccDNA to chromosome length was the highest, reaching 0.13‰ (Fig. 3c). Most of the eccDNA of Arabidopsis thaliana comes from chromosome 2 and mitochondria. Although the length of mitochondria is the shortest, the eccDNA produced by shedding from mitochondria accounts for about one-third of the total (Fig. 3d). The eccDNA of Mus musculus is mainly derived from autosomes. Through comparison, it is not difficult to find that

the source distribution of eccDNA on chromosomes of different species is diverse.

Utility and discussion

Web interface

To ensure that users can use these datasets of eccDNA which we collected and processed, we have developed an online website for browsing and querying the information. The whole website is divided into home, search, blast, statistics, contribute, download, and about.

Search

On the search page, a user can search and browse the basic information sets of eccDNA molecule (Fig. 4a and b). Users can select species or enter ID, gene, chromosome and sample type to retrieve eccDNA directionally. The retrieved eccDNA will be displayed at the bottom of the page. We have also placed examples for users' reference by class under the input box of retrieval. Of course, users can directly view the information of eccDNAs related to a species by clicking the button in "Organism Browse", and the information includes the ID of the eccDNA in the TeCD database, the chromosome to which it belongs, start and end sites in the chromosome, length of the sequence, nucleotide arrangement of the sequence (Fig. 4a). By clicking the ID of the eccDNA,

it would jump to a detailed information page of the eccDNA. The detail page provides more information about the eccDNA (Fig. 4b). Users can find the notes of each row of information on the "About" page.

The information of genes covered by a retrieval eccDNA can be obtained by clicking the gene symbol and jumping to detailed page of gene in NCBI, and detailed function of the gene can be shown in the NCBI page. Each transposable element of eccDNA is linked to a visualization page of Dfam (<https://www.dfam.org/>) which is a eukaryotic transposable element database [40, 41]. It can visualize the distribution of transposable sub-components on a sequence, provide a detailed interpretation of components and download the data. And the retrieval data of Dfam is highly matched with the data in TeCD such as eccDNA sites. The sequencing approach of each eccDNA is also documented on the detail page. We also provide the source literature of each eccDNA, and user can use the link of PMID to obtain the literature in PubMed database (<https://pubmed.ncbi.nlm.nih.gov/>) (Fig. 4b). Those links enhance the interoperability of data resources and improves the convenience of user retrieval.

BLAST

BLAST (basic local alignment search tool) is a set of analysis tools for similarity comparison in protein or DNA

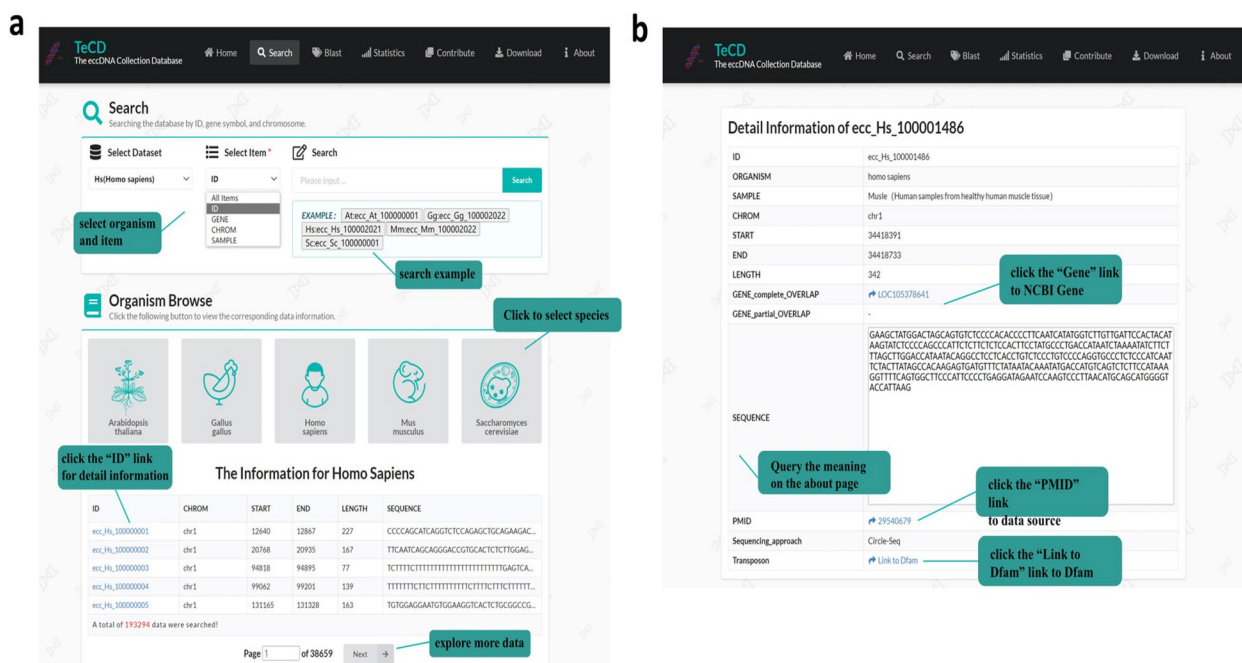


Fig. 4 Screenshot of search page from the TeCD. **a** Search and browse pages. Users can search by organization and item, and we also provide some examples. Selected entry information will be displayed at the bottom of the page for users to browse. **b** Details of an eccDNA. Here, we provide a lot of information about eccDNA, including complete sequence, contained genes, transposons, data sources, etc. All information categories can be queried and explained on the about page. **c** JBrowse style visualization of individual eccDNAs and their location on the host genome

databases [42]. There are three BLAST tools to be provided in alignment in TeCD: BLASTN, TBLASTN, and TBLASTX. The BLASTN can compare and aligns an input nucleic acid sequence, and return the most similar nucleic acid sequence in the TeCD [43, 44]. TBLASTN translates all nucleic acid sequences in TeCD into protein sequences and then align an input protein sequence with the translated sequences [45, 46]. TBLASTX can align a translated input nucleic acid sequence with all translated nucleic acid sequences in TeCD and return the most similar sequences [46]. Users can choose the tool according to sequence information and needs (Fig. 5a).

Statistics

The statistics page presents the aggregated data in TeCD by tables. It includes the number of genes covered on the eccDNA of each species, percentage of eccDNA with complete gene overlap and partial gene overlap among species, the distribution of eccDNAs across species and sample types. Moreover, the distribution of eccDNAs on chromosomes of each species and the sequencing approach of eccDNAs are also summarized.

Contribute, download and about

To continuously expand and improve the database, we have built a submission and upload interface, and

allow users to upload new eccDNA sequences into the TeCD database (Fig. 6a). We will process the upload eccDNA sequences and integrate them into TeCD in the background.

All data in TeCD are shared under the knowledge sharing license agreement, and we provide a download interface for users to download the data from TeCD (Fig. 6c). Of course, if users want to directly download part of the data they retrieved, they can also check download on the search page. As shown in Fig. 6b, by searching eccDNA data on the human chromosome 21, users can manually select some or all of the data to export. To facilitate users to understand and use our database, we have developed an “About” page to query the meaning of specific items in the detailed sequence information in time.

Analyze database sequences using the BLAST

In the *Saccharomyces cerevisiae* data from the TeCD, there were experimental groups with the DNA damaging agent zeocin and control groups without the DNA damaging agent, to explore the effects of DNA damaging agents on eccDNA. As shown in Fig. 7, the control group had an average of 147 eccDNAs per group (181, 42, 248, and 118 eccDNAs); whereas the group with the DNA damaging agent zeocin had an average of 219 eccDNAs (159, 210, 218, and 288 eccDNAs). Thus, the

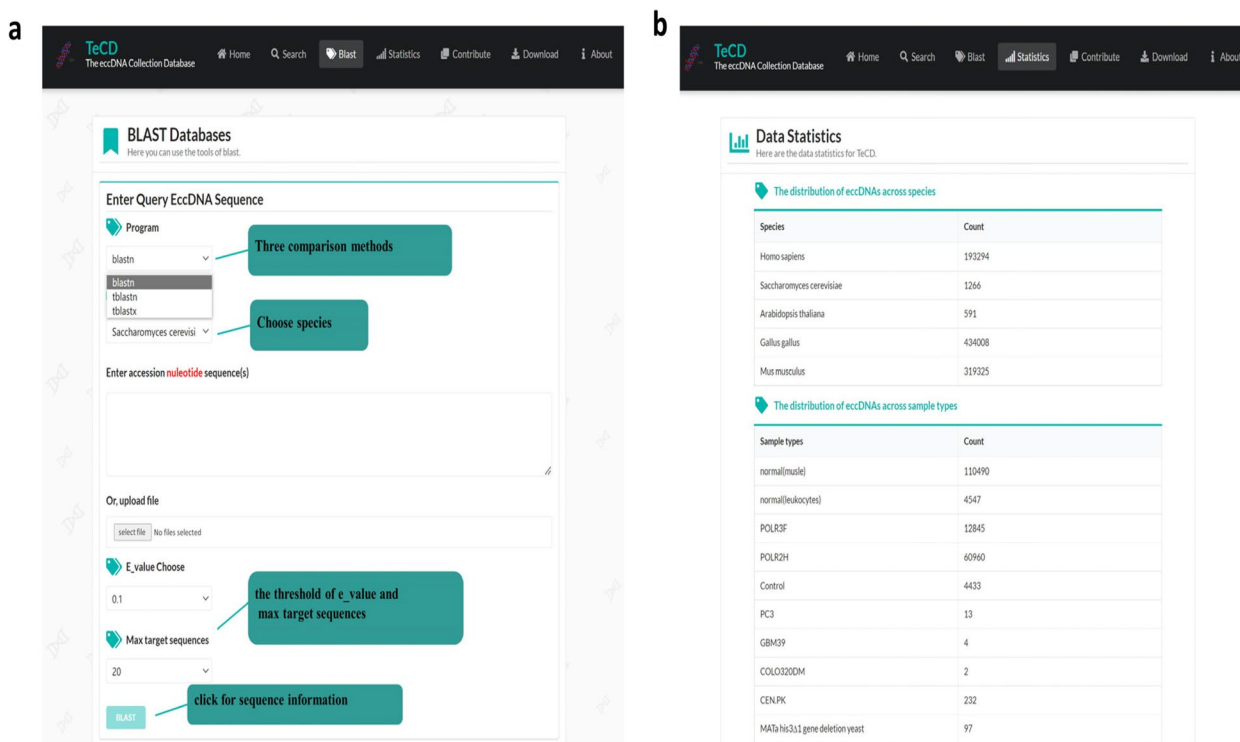


Fig. 5 Screenshot of BLAST and statistics page. **a** Screenshot of blast page from the TeCD. Users can use BLAST to search the database by sequence alignment. **b** Screenshot of statistics page from the TeCD

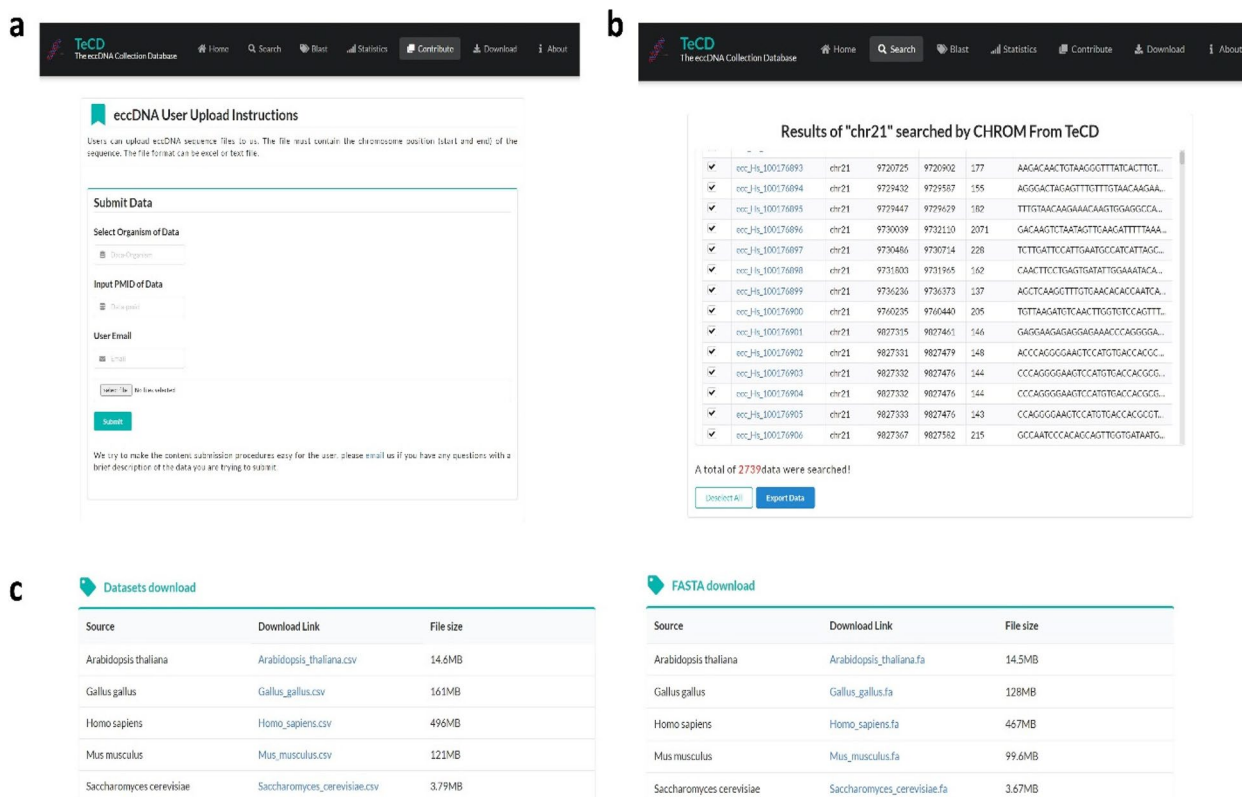


Fig. 6 Screenshot of contribute and download page. **a** Screenshot of contribute page from the TeCD. Users can export some data optionally. **c** Screenshot of download page from the TeCD. TeCD provides download links of all data by species, and the data formats are .csv and .fa

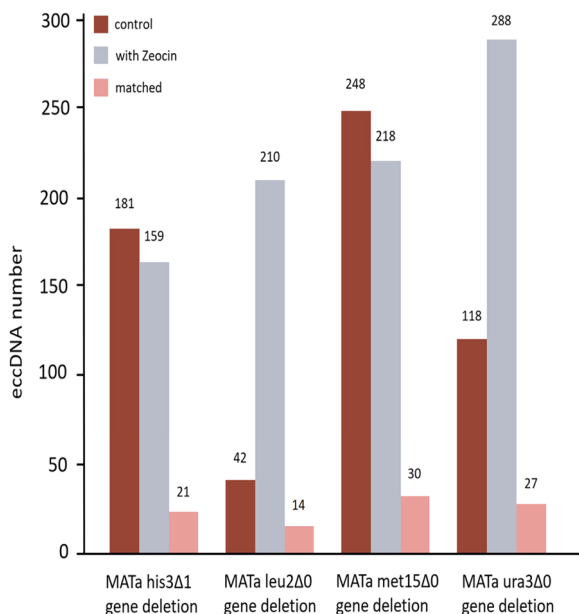


Fig. 7 The number of eccDNA with and without DNA damaging. Taking *Saccharomyces cerevisiae* as an example, a bidirectional comparison of nucleic acid sequences was performed by BLAST

samples treated with the DNA damaging agent zeocin (219 eccDNAs) showed more eccDNA than untreated samples (147 eccDNAs). We further processed the data and performed bidirectional comparisons of eccDNA sequences obtained from each experimental group and its control group using the platform-built BLAST tool. The result showed that the eccDNA matched by bidirectional comparisons in each group to account for 11%-13% of the average eccDNA in the control and experimental groups. Thus, this result can not only verify that the DNA damaging agent Zeocin can damage the single-strand and double-strand DNA of cells to form eccDNA [23], but also can be used to evaluate the level of DNA strand breakage induced by the DNA damaging agent Zeocin and its effect on eccDNA formation.

Conclusion

To provide the chromosome source, complete base sequence, and related gene information of eccDNA in different species, we have carried out a large amount of data collection and sorting out, and developed TeCD. We obtained the site information of eccDNA

from published literature and matched it with reference genome and gene library data. Thus, we obtained the partially overlapped genes, complete overlapped genes, and transposons. In addition, the sequencing approaches for obtaining eccDNA were also collated and recorded. TeCD web interface allows users to query on-demand and browse the details of each eccDNA. To facilitate users to compare and identify sequences by themselves, we have provided the BLAST function on the web page and all the processed data in TeCD can be downloaded directly by users. TeCD will be updated regularly according to the latest sequencing data and data submitted by users as an interactive website. Based on our knowledge, TeCD is the first database containing the sequence and genomic information of extrachromosomal circular DNA of eukaryotes and BLAST sequence alignment function. Anyway, TeCD will become a powerful tool for researchers to browse and retrieve eccDNA data and further explore the biological functions and contributions of eccDNA.

Abbreviations

EccDNA	Extrachromosomal circular DNA
TeCD	The eccDNA Collection Database
BLAST	Basic Local Alignment Search Tool
NCBI	National Center for Biotechnology Information
ID	Identifier
TAIR	The Arabidopsis Information Resource
HEK	Human Embryonic Kidney
POLR2H	RNA Polymerase II, I And III Subunit H
POLR3F	RNA Polymerase III Subunit F
PC	Prostate Cancer
COLO	Colon Cancer
GBM	Glioblastoma
PMID	PubMed Unique Identifier

Acknowledgements

Not applicable.

Authors' contributions

JG: collection of data, analysis and interpretation of data, writing of the manuscript. ZZ: construction and maintenance of the database. QL: collection of data. XC and XL: conception and design of the database, manuscript revision. All authors read and approved the final manuscript.

Funding

This work was supported by grants from Zhejiang Provincial Natural Science Foundation of China (No. LZ22C060001), the research funds of Hangzhou Institute for advanced study, UCAS (No. 2022ZZ01013), the Key Project of Natural Science of Anhui Provincial Education Department (No. KJ2020A0018), the Key project of Anhui Finance and Economics University (No. ackyb20015), and Project of teaching and research of Department of Education of Anhui Province (No.2020xxsxxk014).

Availability of data and materials

All the data are free to use for academic purpose at <http://www.eccdna.org:2022> or <http://122.224.251.240:2022>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

Not applicable.

Received: 4 May 2022 Accepted: 13 January 2023

Published online: 27 January 2023

References

- Koche RP, Rodriguez-Fos E, Helmsauer K, Burkert M, MacArthur IC, Maag J, Chamorro R, Munoz-Perez N, Puiggros M, Dorado Garcia H, et al. Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma. *Nat Genet.* 2020;52(1):29–34. <https://doi.org/10.1038/s41588-019-0547-z>.
- Yan Y, Guo G, Huang J, Gao M, Zhu Q, Zeng S, Gong Z, Xu Z. Current understanding of extrachromosomal circular DNA in cancer pathogenesis and therapeutic resistance. *J Hematol Oncol.* 2020;13(1):124. <https://doi.org/10.1186/s13045-020-00960-9>.
- Turner KM, Deshpande V, Beyter D, Koga T, Rusert J, Lee C, Li B, Arden K, Ren B, Nathanson DA, et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature.* 2017;543(7643):122–5. <https://doi.org/10.1038/nature21356>.
- Mansidor A, Molinar T, Srivastava P, Dartis DD, Delgado AP, Blitzblau HG, Klein H, Hochwagen A. Genomic Copy-Number Loss Is Rescued by Self-Limiting Production of DNA Circles. *Mol Cell.* 2018;72(3):583. <https://doi.org/10.1016/j.molcel.2018.08.036>.
- Hull RM, Houseley J. The adaptive potential of circular DNA accumulation in ageing cells. *Curr Genet.* 2020;66(5):889–94. <https://doi.org/10.1007/s00294-020-01069-9>.
- Paulsen T, Kumar P, Koseoglu MM, Dutta A. Discoveries of Extrachromosomal Circles of DNA in Normal and Tumor Cells. *Trends Genet.* 2018;34(4):270–8. <https://doi.org/10.1016/j.tig.2017.12.010>.
- Kumar P, Dillon LW, Shibata Y, Jazaeri AA, Jones DR, Dutta A. Normal and Cancerous Tissues Release Extrachromosomal Circular DNA (eccDNA) into the Circulation. *Mol Cancer Res.* 2017;15(9):1197–205. <https://doi.org/10.1158/1541-7786.Mcr-17-0095>.
- Ling X, Han Y, Meng J, Zhong B, Chen J, Zhang H, Qin J, Pang J, Liu L. Small extrachromosomal circular DNA (eccDNA): major functions in evolution and cancer. *Mol Cancer.* 2021;20(1):113. <https://doi.org/10.1186/s12943-021-01413-8>.
- Wang M, Chen X, Yu F, Ding H, Zhang Y, Wang K. Extrachromosomal Circular DNAs: Origin, formation and emerging function in Cancer. *Int J Biol Sci.* 2021;17(4):1010–25. <https://doi.org/10.7150/ijbs.54614>.
- Wang T, Zhang H, Zhou Y, Shi J. Extrachromosomal circular DNA: a new potential role in cancer progression. *J Transl Med.* 2021;19(1):257. <https://doi.org/10.1186/s12967-021-02927-x>.
- Cohen S, Mechali M. A novel cell-free system reveals a mechanism of circular DNA formation from tandem repeats. *Nucleic Acids Res.* 2001;29(12):2542–8. <https://doi.org/10.1093/nar/29.12.2542>.
- Gaubatz JW. Extrachromosomal circular DNAs and genomic sequence plasticity in eukaryotic cells. *Mutat Res.* 1990;237(5–6):271–92. [https://doi.org/10.1016/0921-8734\(90\)90009-g](https://doi.org/10.1016/0921-8734(90)90009-g).
- Motejlek K, Schindler D, Assum G, Krone W. Increased amount and contour length distribution of small polydisperse circular DNA (spcDNA) in Fanconi anemia. *Mutat Res.* 1993;293(3):205–14. [https://doi.org/10.1016/0921-8777\(93\)90071-n](https://doi.org/10.1016/0921-8777(93)90071-n).
- Peng L, Zhou N, Zhang CY, Li GC, Yuan XQ. eccDNAdb: a database of extrachromosomal circular DNA profiles in human cancers. *Oncogene.* 2022;41(19):2696–705. <https://doi.org/10.1038/s41388-022-02286-x>.
- Zhao XL, Shi LS, Ruan SS, Bi WJ, Chen YF, Chen L, Liu YF, Li MK, Qiao J, Mao FB. CircleBase: an integrated resource and analysis platform for human eccDNAs. *Nucleic Acids Res.* 2022;50(D1):D72–82. <https://doi.org/10.1093/nar/gkab1104>.
- Cohen S, Segal D. Extrachromosomal circular DNA in eukaryotes: possible involvement in the plasticity of tandem repeats. *Cytogenet Genome Res.* 2009;124(3–4):327–38. <https://doi.org/10.1159/000218136>.

17. Hull RM, King M, Pizza G, Krueger F, Vergara X, Houseley J. Transcription-induced formation of extrachromosomal DNA during yeast ageing. *PLoS Biol.* 2019;17(12):e3000471. <https://doi.org/10.1371/journal.pbio.3000471>.
18. Moller HD. Circle-Seq: Isolation and Sequencing of Chromosome-Derived Circular DNA Elements in Cells. *Methods Mol Biol.* 2020;2119:165–81. https://doi.org/10.1007/978-1-0716-0323-9_15.
19. Kumar P, Kiran S, Saha S, Su Z, Paulsen T, Chatrath A, Shibata Y, Shibata E, Dutta A. ATAC-seq identifies thousands of extrachromosomal circular DNA in cancer and cell lines. *Sci Adv.* 2020;6(20):eaba2489. <https://doi.org/10.1126/sciadv.aba2489>.
20. Moller HD, Mohiyuddin M, Prada-Luengo I, Sailani MR, Halling JF, Plomgaard P, Maretty L, Hansen AJ, Snyder MP, Pilegaard H, et al. Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. *Nat Commun.* 2018;9(1):1069. <https://doi.org/10.1038/s41467-018-03369-8>.
21. Moller HD, Bojsen RK, Tachibana C, Parsons L, Botstein D, Regenberg B. Genome-wide Purification of Extrachromosomal Circular DNA from Eukaryotic Cells. *J Vis Exp.* 2016;110:e54239. <https://doi.org/10.3791/54239>.
22. Paulsen T, Shibata Y, Kumar P, Dillon L, Dutta A. Small extrachromosomal circular DNAs, microDNA, produce short regulatory RNAs that suppress gene expression independent of canonical promoters. *Nucleic Acids Res.* 2019;47(9):4586–96. <https://doi.org/10.1093/nar/gkz155>.
23. Moller HD, Parsons L, Jorgensen TS, Botstein D, Regenberg B. Extrachromosomal circular DNA is common in yeast. *P Natl Acad Sci USA.* 2015;112(24):E3114–22. <https://doi.org/10.1073/pnas.1508825112>.
24. Wu S, Turner KM, Nguyen N, Raviram R, Erb M, Santini J, Luebeck J, Rajkumar U, Diao Y, Li B, et al. Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature.* 2019;575(7784):699–703. <https://doi.org/10.1038/s41586-019-1763-5>.
25. Mehta D, Cornet L, Hirsch-Hoffmann M, Zaidi SS, Vanderschuren H. Full-length sequencing of circular DNA viruses and extrachromosomal circular DNA using CIDER-Seq. *Nat Protoc.* 2020;15(5):1673–89. <https://doi.org/10.1038/s41596-020-0301-0>.
26. Morton AR, Dogan-Artun N, Faber ZJ, MacLeod G, Bartels CF, Piazza MS, Allan KC, Mack SC, Wang X, Gimple RC, et al. Functional Enhancers Shape Extrachromosomal Oncogene Amplifications. *Cell.* 2019;179(6):1330–1341 e1313. <https://doi.org/10.1016/j.cell.2019.10.039>.
27. Sin STK, Jiang P, Deng J, Ji L, Cheng SH, Dutta A, Leung TY, Chan KCA, Chiu RWK, Lo YMD. Identification and characterization of extrachromosomal circular DNA in maternal plasma. *Proc Natl Acad Sci U S A.* 2020;117(3):1658–65. <https://doi.org/10.1073/pnas.1914949117>.
28. Shoura MJ, Gabdank I, Hansen L, Merker J, Gotlib J, Levene SD, Fire AZ. Intricate and Cell Type-Specific Populations of Endogenous Circular DNA (eccDNA) in *Caenorhabditis elegans* and *Homo sapiens*. *G3 (Bethesda).* 2017;7(10):3295–303. <https://doi.org/10.1534/g3.117.300141>.
29. Moller HD, Ramos-Madriral J, Prada-Luengo I, Gilbert MTP, Regenberg B. Near-Random Distribution of Chromosome-Derived Circular DNA in the Condensed Genome of Pigeons and the Larger, More Repeat-Rich Human Genome. *Genome Biol Evol.* 2020;12(1):3762–77. <https://doi.org/10.1093/gbe/evz281>.
30. van de Werken HJ, Landan G, Holwerda SJ, Hoichman M, Klous P, Chachik R, Splinter E, Valdes-Quezada C, Oz Y, Bouwman BA, et al. Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat Methods.* 2012;9(10):969–72. <https://doi.org/10.1038/nmeth.2173>.
31. Epstein L, Hunter JC, Arwady MA, Tsai V, Stein L, Gribogiannis M, Frias M, Guh AY, Laufer AS, Black S, et al. New Delhi metallo-beta-lactamase-producing carbapenem-resistant *Escherichia coli* associated with exposure to duodenoscopes. *JAMA.* 2014;312(14):1447–55. <https://doi.org/10.1001/jama.2014.12720>.
32. Grzywacz B, Chobanov DP, Maryanska-Nadachowska A, Karamysheva TV, Heller KG, Warchalowska-Sliwa E. A comparative study of genome organization and inferences for the systematics of two large bushcricket genera of the tribe Barbitistini (Orthoptera: Tettigoniidae: Phaneropterinae). *Bmc Evol Biol.* 2014;14:48. <https://doi.org/10.1186/1471-2148-14-48>.
33. Molin WT, Yaguchi A, Blenner M, Sasaki CA. The EccDNA Replicon: A Heritable, Extranuclear Vehicle That Enables Gene Amplification and Glyphosate Resistance in *Amaranthus palmeri*. *Plant Cell.* 2020;32(7):2132–40. <https://doi.org/10.1105/tpc.20.00099>.
34. Wang KY, Tian H, Wang LQ, Wang L, Tan YC, Zhang ZT, Sun K, Yin M, Wei QG, Guo BH, et al. Deciphering extrachromosomal circular DNA in Arabidopsis. *Comput Struct Biotech.* 2021;19:1176–83. <https://doi.org/10.1016/j.csbj.2021.01.043>.
35. Dillon LW, Kumar P, Shibata Y, Wang YH, Willcox S, Griffith JD, Pommier Y, Takeda S, Dutta A. Production of Extrachromosomal MicroDNAs Is Linked to Mismatch Repair Pathways and Transcriptional Activity. *Cell Rep.* 2015;11(11):1749–59. <https://doi.org/10.1016/j.celrep.2015.05.020>.
36. Boratyn GM, Camacho C, Cooper PS, Coulouris G, Fong A, Ma N, Madden TL, Matten WT, McGinnis SD, Merezhuk Y, et al. BLAST: a more efficient report with usability improvements. *Nucleic Acids Res.* 2013;41(Web Server issue):W29–33. <https://doi.org/10.1093/nar/gkt282>.
37. Garcia-Hernandez M, Berardini TZ, Chen G, Crist D, Doyle A, Huala E, Knee E, Lambrecht M, Miller N, Mueller LA, et al. TAIR: a resource for integrated Arabidopsis data. *Funct Integr Genomics.* 2002;2(6):239–53. <https://doi.org/10.1007/s10142-002-0077-z>.
38. Rhee SY, Beavis W, Berardini TZ, Chen GH, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M, et al. The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. *Nucleic Acids Res.* 2003;31(1):224–8. <https://doi.org/10.1093/nar/gkg076>.
39. Haber JE. Mating-type genes and MAT switching in *Saccharomyces cerevisiae*. *Genetics.* 2012;191(1):33–64. <https://doi.org/10.1534/genetics.111.134577>.
40. Wheeler TJ, Clements J, Eddy SR, Hubley R, Jones TA, Jurka J, Smit AFA, Finn RD. Dfam: a database of repetitive DNA based on profile hidden Markov models. *Nucleic Acids Res.* 2013;41(D1):D70–82. <https://doi.org/10.1093/nar/gks1265>.
41. Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mobile DNA-Uk.* 2021;12(1):2. <https://doi.org/10.1186/s13100-020-00230-y>.
42. McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* 2004;32:W20–5. <https://doi.org/10.1093/nar/gkh435>.
43. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res.* 2008;36(Web Server issue):W5–9. <https://doi.org/10.1093/nar/gkn201>.
44. Ladunga I. Finding Similar Nucleotide Sequences Using Network BLAST Searches. *Curr Protoc Bioinformatics.* 2017;58:3 3 1–3 25. <https://doi.org/10.1002/cpbi.29>.
45. Zhang ZY, Zhang HM, Li DS, Xiong TY, Fang SG. Characterization of the beta-defensin genes in giant panda. *Sci Rep.* 2018;8(1):12308. <https://doi.org/10.1038/s41598-018-29898-2>.
46. Zhang X, Mao Y, Huang Z, Qu M, Chen J, Ding S, Hong J, Sun T. Transcriptome analysis of the Octopus vulgaris central nervous system. *PLoS One.* 2012;7(6):e40320. <https://doi.org/10.1371/journal.pone.0040320>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

