DDBJ Updates in 2024

Yaz Nakamura
中村保一

The 9th Big Data Forum for Life and Health Sciences & The 21st Asian Bioinformatics Consortium Symposium
Oct.. 16-18, 2024 Beijing

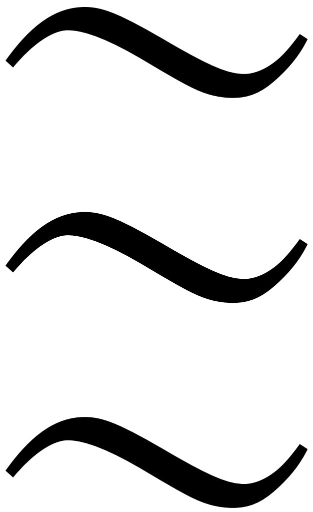
Center Village Keep Number one

中村保一

Nakamura Yasukazu

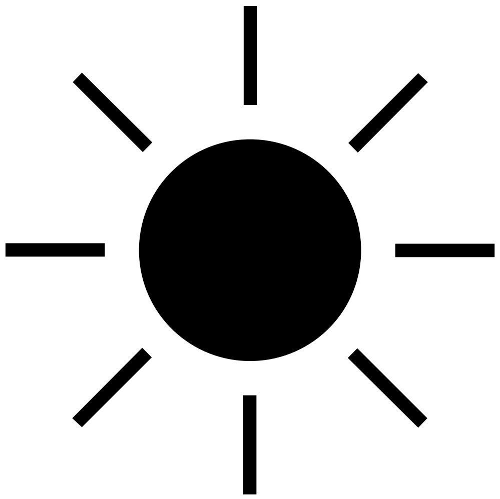
mountain

山



river

川

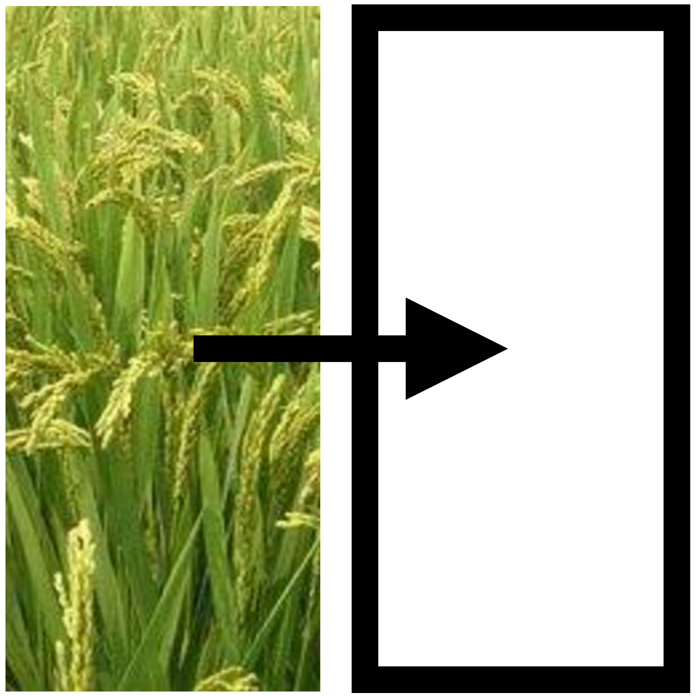


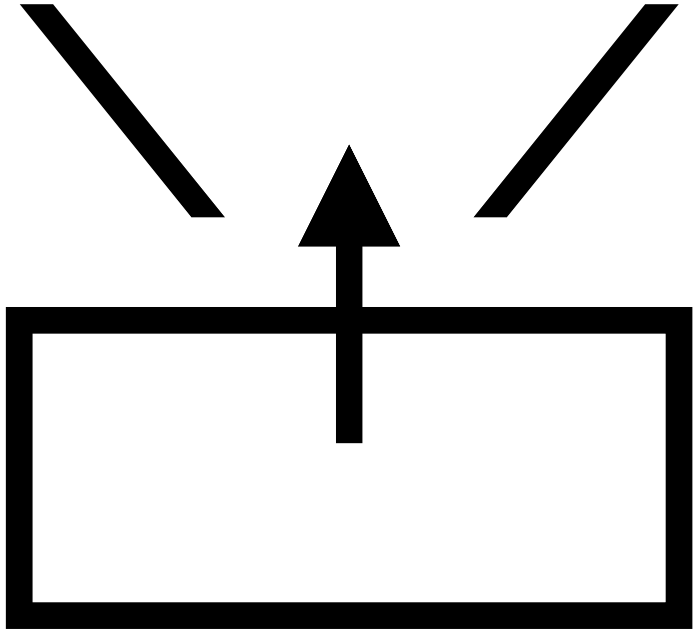
sun/day

日

私 INSDC 公

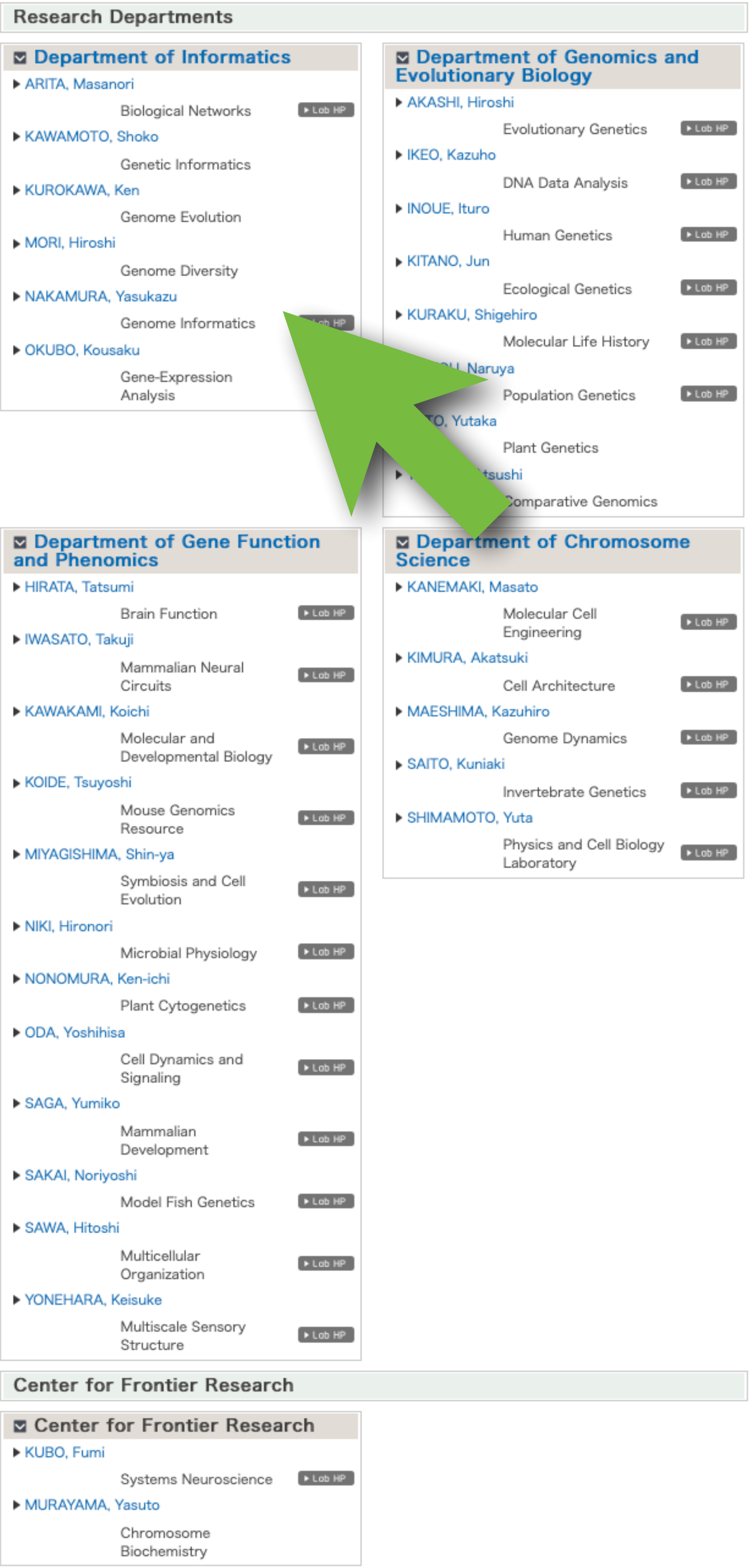
private → public





National Institute of Genetics

National Institute of Genetics: www.nig.ac.jp



Research Departments

☑ Department of Informatics
- ARITA, Masanori
  Biological Networks ▸ Lab HP
- KAWAMOTO, Shoko
  Genetic Informatics
- KUROKAWA, Ken
  Genome Evolution
- MORI, Hiroshi
  Genome Diversity
- NAKAMURA, Yasukazu
  Genome Informatics ▸ Lab HP
- OKUBO, Kousaku
  Gene-Expression Analysis

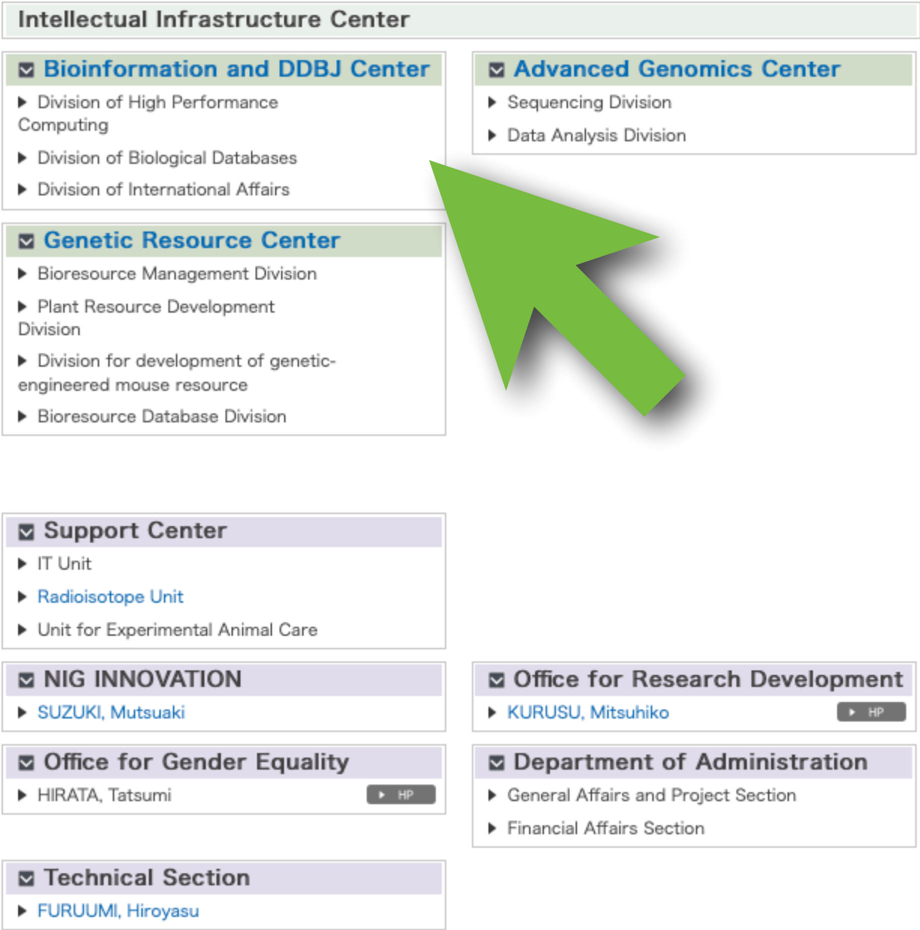
☑ Department of Genomics and Evolutionary Biology
- AKASHI, Hiroshi
  Evolutionary Genetics ▸ Lab HP
- IKEO, Kazuho
  DNA Data Analysis ▸ Lab HP
- INOUE, Ituro
  Human Genetics ▸ Lab HP
- KITANO, Jun
  Ecological Genetics ▸ Lab HP
- KURAKU, Shigehiro
  Molecular Life History ▸ Lab HP
- U, Naruya
  Population Genetics ▸ Lab HP
- TO, Yutaka
  Plant Genetics
- tsushi
  Comparative Genomics

☑ Department of Gene Function and Phenomics
- HIRATA, Tatsumi
  Brain Function ▸ Lab HP
- IWASATO, Takuji
  Mammalian Neural Circuits ▸ Lab HP
- KAWAKAMI, Koichi
  Molecular and Developmental Biology ▸ Lab HP
- KOIDE, Tsuyoshi
  Mouse Genomics Resource ▸ Lab HP
- MIYAGISHIMA, Shin-ya
  Symbiosis and Cell Evolution ▸ Lab HP
- NIKI, Hironori
  Microbial Physiology ▸ Lab HP
- NONOMURA, Ken-ichi
  Plant Cytogenetics ▸ Lab HP
- ODA, Yoshihisa
  Cell Dynamics and Signaling ▸ Lab HP
- SAGA, Yumiko
  Mammalian Development ▸ Lab HP
- SAKAI, Noriyoshi
  Model Fish Genetics ▸ Lab HP
- SAWA, Hitoshi
  Multicellular Organization ▸ Lab HP
- YONEHARA, Keisuke
  Multiscale Sensory Structure ▸ Lab HP

☑ Department of Chromosome Science
- KANEMAKI, Masato
  Molecular Cell Engineering ▸ Lab HP
- KIMURA, Akatsuki
  Cell Architecture ▸ Lab HP
- MAESHIMA, Kazuhiro
  Genome Dynamics ▸ Lab HP
- SAITO, Kuniaki
  Invertebrate Genetics ▸ Lab HP
- SHIMAMOTO, Yuta
  Physics and Cell Biology Laboratory ▸ Lab HP

Center for Frontier Research

☑ Center for Frontier Research
- KUBO, Fumi
  Systems Neuroscience ▸ Lab HP
- MURAYAMA, Yasuto
  Chromosome Biochemistry



Intellectual Infrastructure Center

☑ Bioinformation and DDBJ Center
- Division of High Performance Computing
- Division of Biological Databases
- Division of International Affairs

☑ Advanced Genomics Center
- Sequencing Division
- Data Analysis Division

☑ Genetic Resource Center
- Bioresource Management Division
- Plant Resource Development Division
- Division for development of genetic-engineered mouse resource
- Bioresource Database Division

☑ Support Center
- IT Unit
- Radioisotope Unit
- Unit for Experimental Animal Care

☑ NIG INNOVATION
- SUZUKI, Mutsuaki

☑ Office for Research Development
- KURUSU, Mitsuhiko ▸ HP

☑ Office for Gender Equality
- HIRATA, Tatsumi ▸ HP

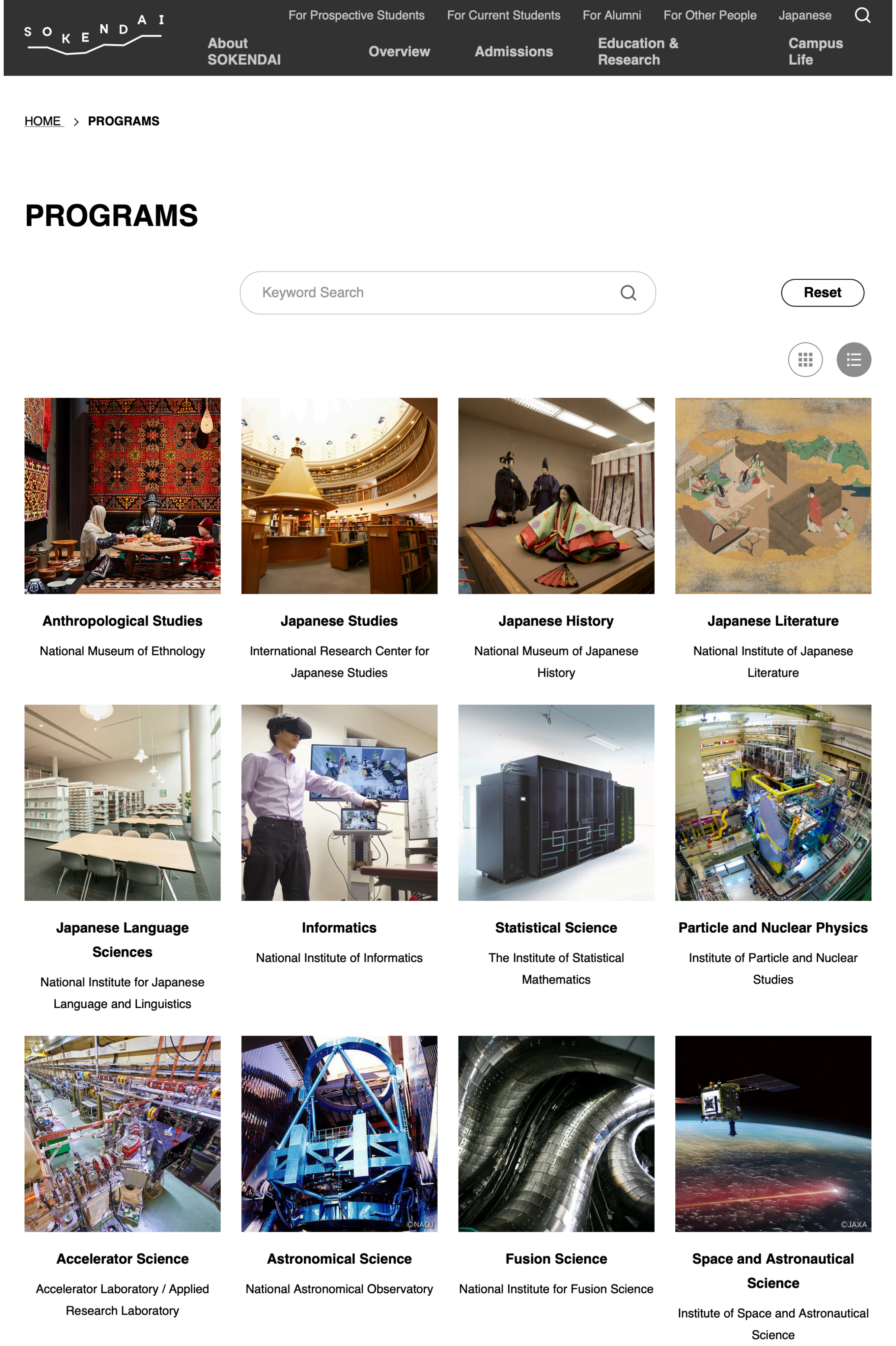
☑ Department of Administration
- General Affairs and Project Section
- Financial Affairs Section

☑ Technical Section
- FURUUMI, Hiroyasu

NATIONAL INSTITUTE OF GENETICS · JAPAN · NIG

SOKENDAI

The Graduate University for Advanced Studies



HOME > PROGRAMS

PROGRAMS

Anthropological Studies
National Museum of Ethnology

Japanese Studies
International Research Center for Japanese Studies

Japanese History
National Museum of Japanese History

Japanese Literature
National Institute of Japanese Literature

Japanese Language Sciences
National Institute for Japanese Language and Linguistics

Informatics
National Institute of Informatics

Statistical Science
The Institute of Statistical Mathematics

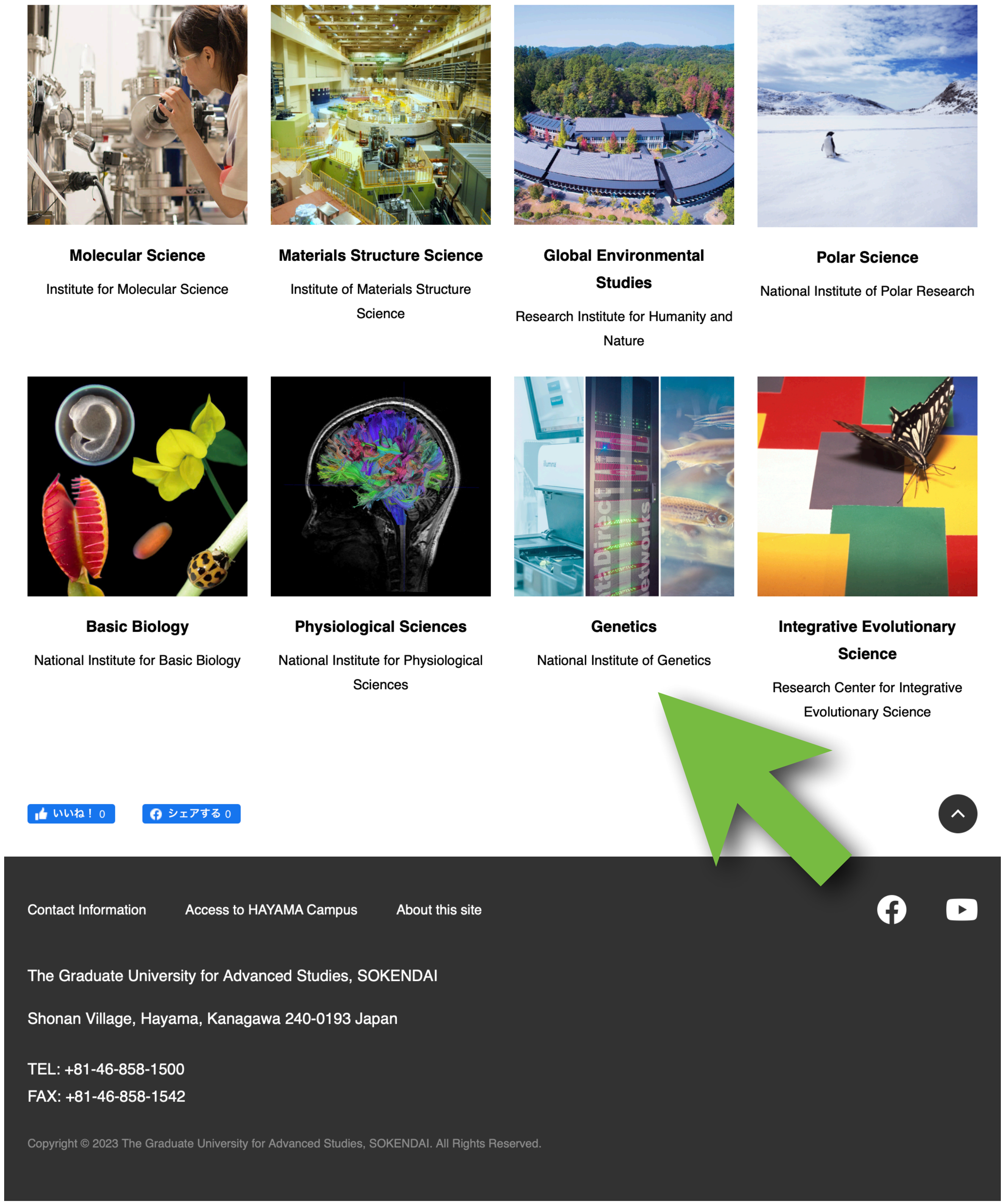
Particle and Nuclear Physics
Institute of Particle and Nuclear Studies

Accelerator Science
Accelerator Laboratory / Applied Research Laboratory

Astronomical Science
National Astronomical Observatory

Fusion Science
National Institute for Fusion Science

Space and Astronautical Science
Institute of Space and Astronautical Science



Molecular Science
Institute for Molecular Science

Materials Structure Science
Institute of Materials Structure Science

Global Environmental Studies
Research Institute for Humanity and Nature

Polar Science
National Institute of Polar Research

Basic Biology
National Institute for Basic Biology

Physiological Sciences
National Institute for Physiological Sciences

Genetics
National Institute of Genetics

Integrative Evolutionary Science
Research Center for Integrative Evolutionary Science

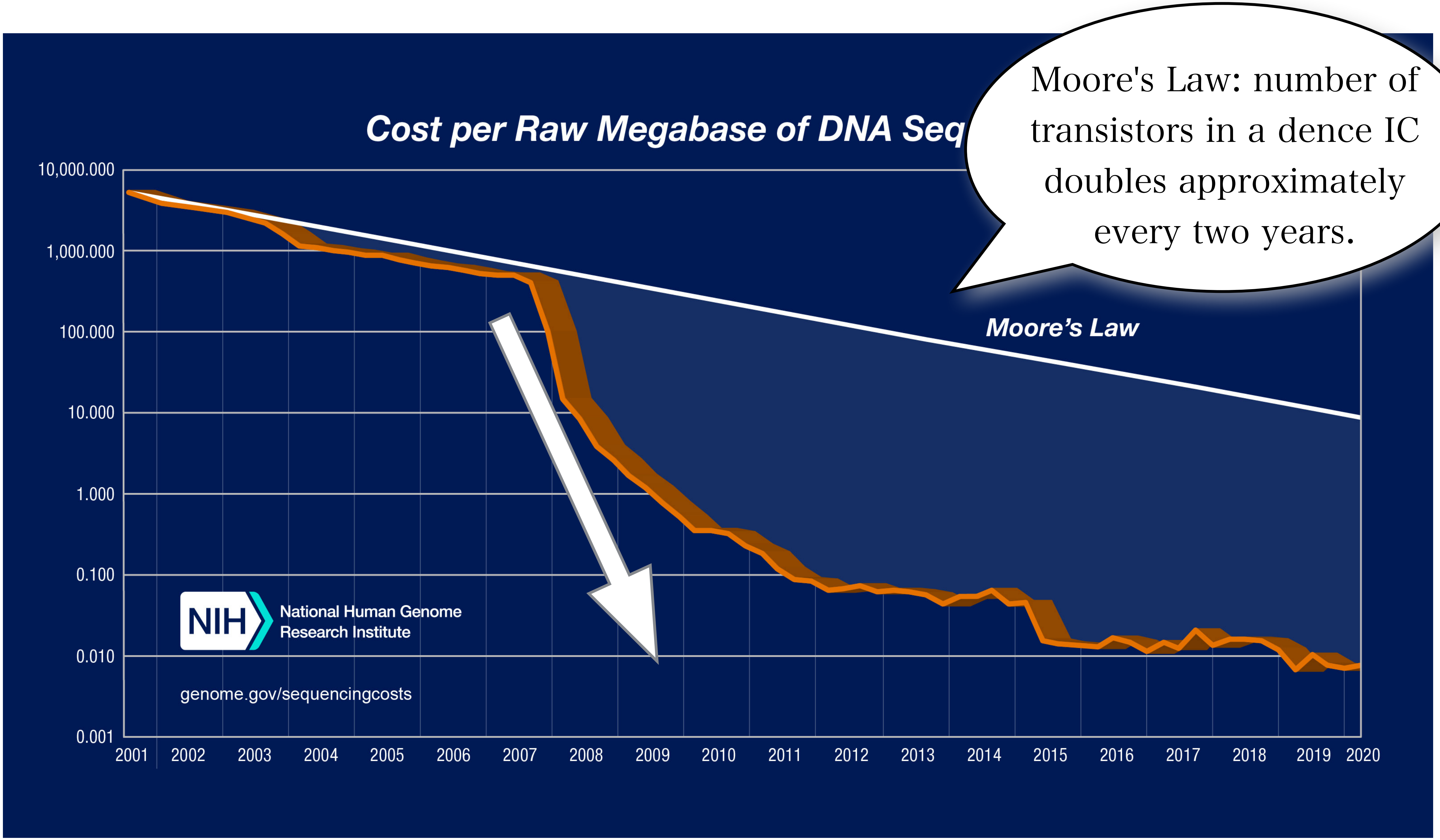
The Graduate University for Advanced Studies, SOKENDAI

Shonan Village, Hayama, Kanagawa 240-0193 Japan

TEL: +81-46-858-1500
FAX: +81-46-858-1542

DDBJ and the INDSC

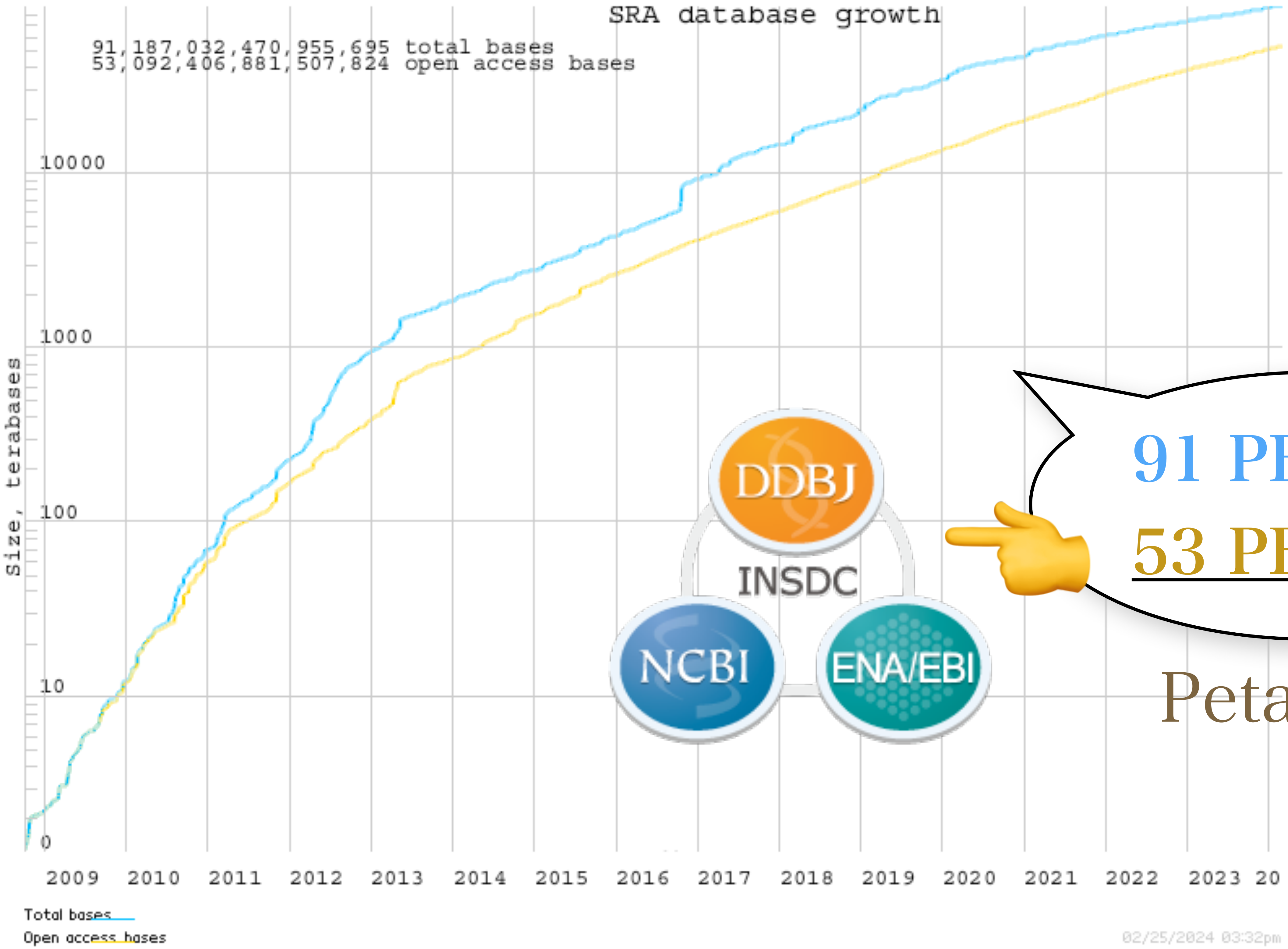
Sequencing cost



genome.gov/sequencingcosts

Moore's Law: number of transistors in a dence IC doubles approximately every two years.

Sequence Read Archive growth (2024.2.25)



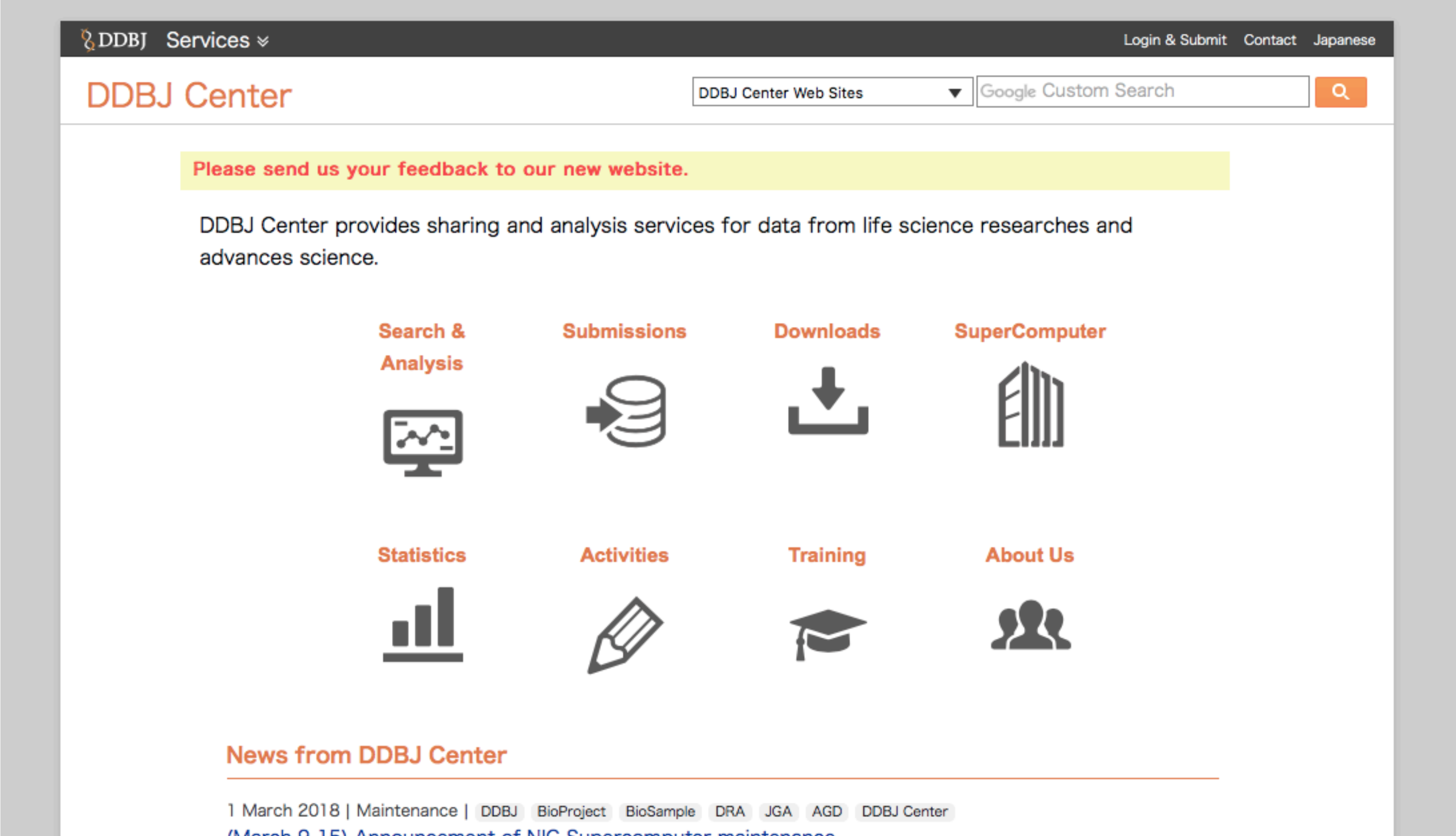
91 PB total

53 PB open

Peta: 10^{15}

DDBJ

DNA DataBank of Japan



DDBJ Center

Please send us your feedback to our new website.

DDBJ Center provides sharing and analysis services for data from life science researches and advances science.

News from DDBJ Center

1 March 2018 | Maintenance | DDBJ BioProject BioSample DRA JGA AGD DDBJ Center

INSDC

The international nucleotide sequence database collaboration

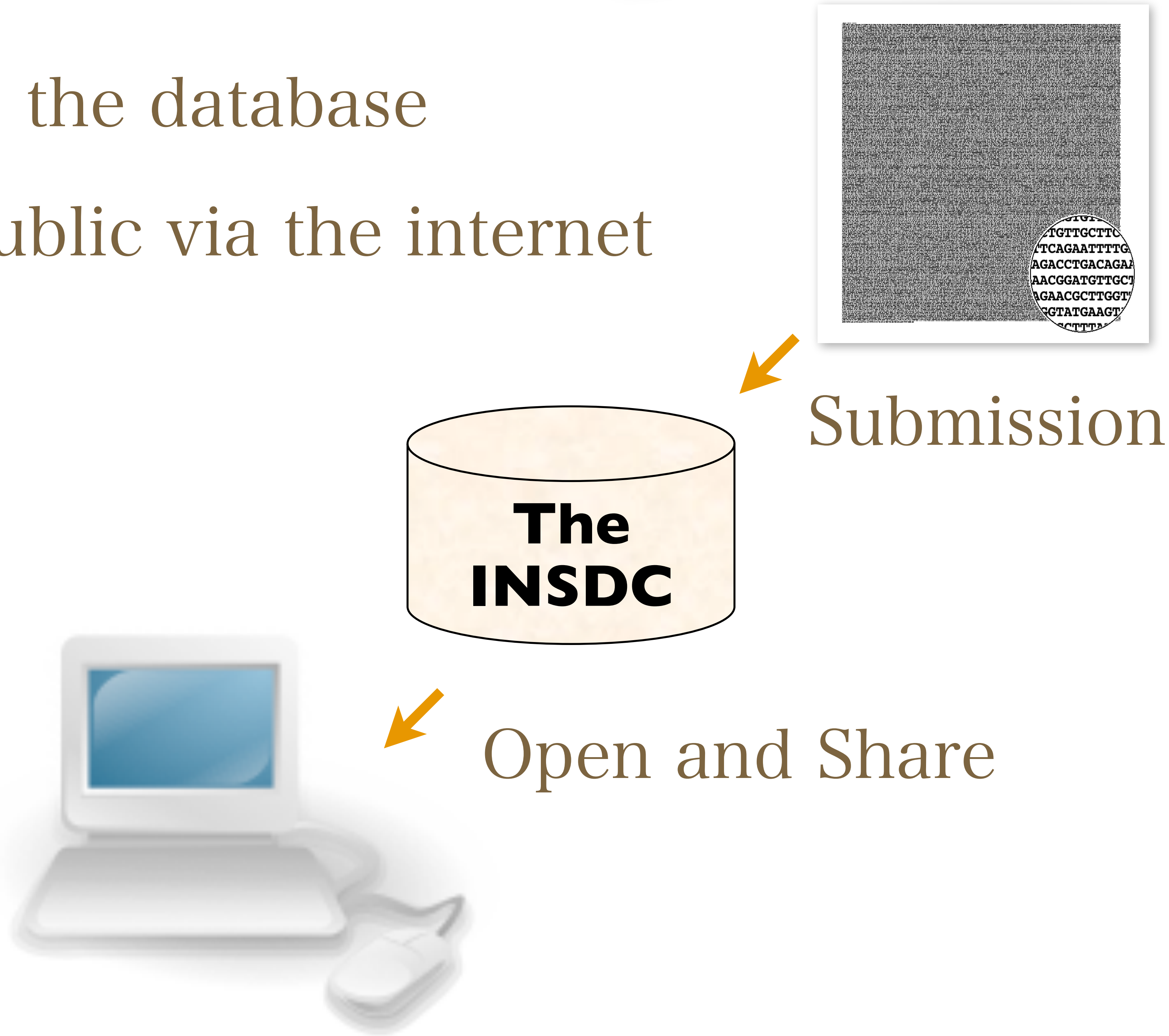
International Nucleotide Sequence Database Collaboration

The International Nucleotide Sequence Database Collaboration (INSDC) archives nucleotide sequence data, from raw to assembled and annotated sequences, from around the world.

The business of DNA Databanks — INSDC

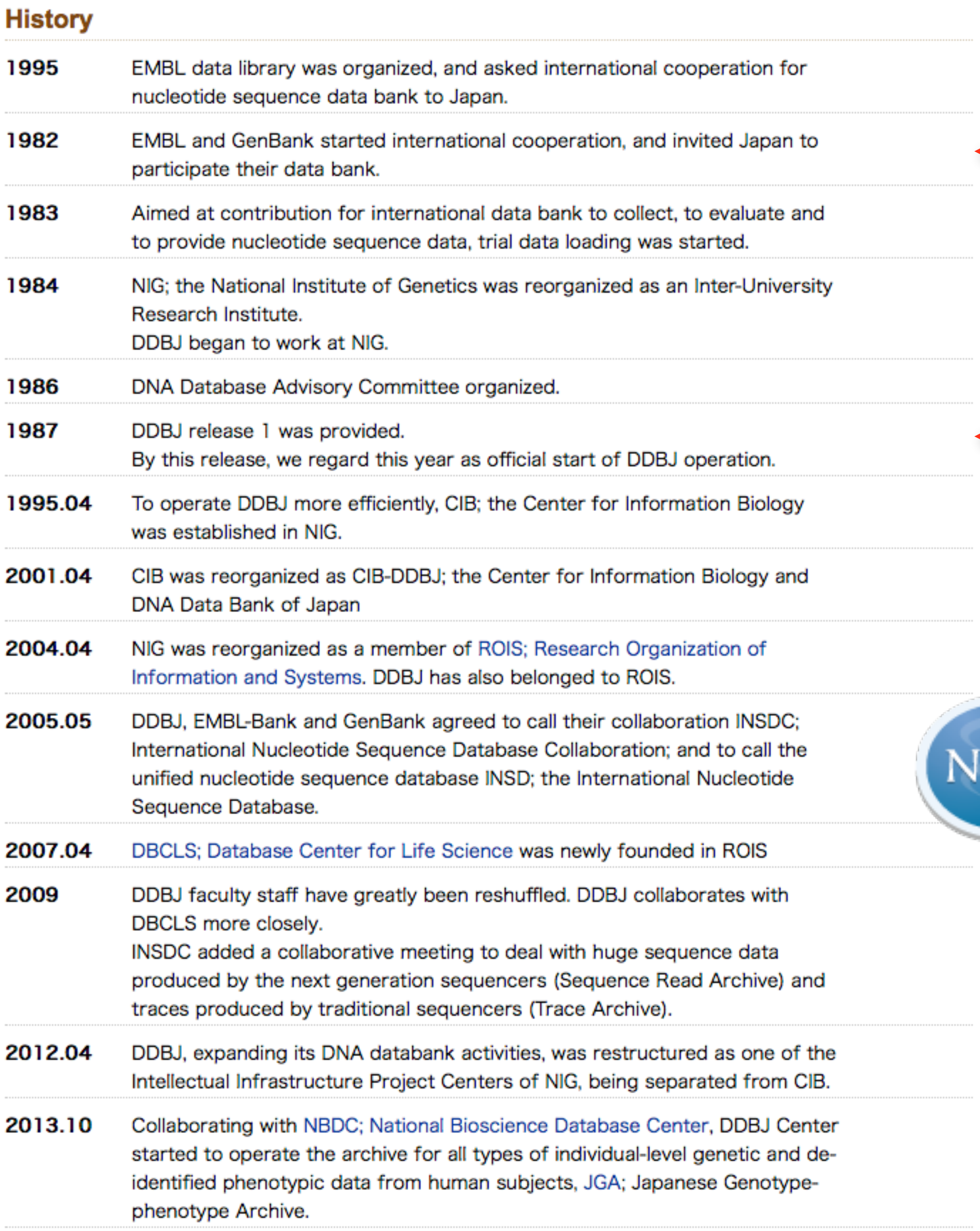
- Submitted Nucleotide Sequence
  - We check its data and metadata
  - We put it into the database
  - We make it public via the internet

Curation!



Submission

Open and Share



History

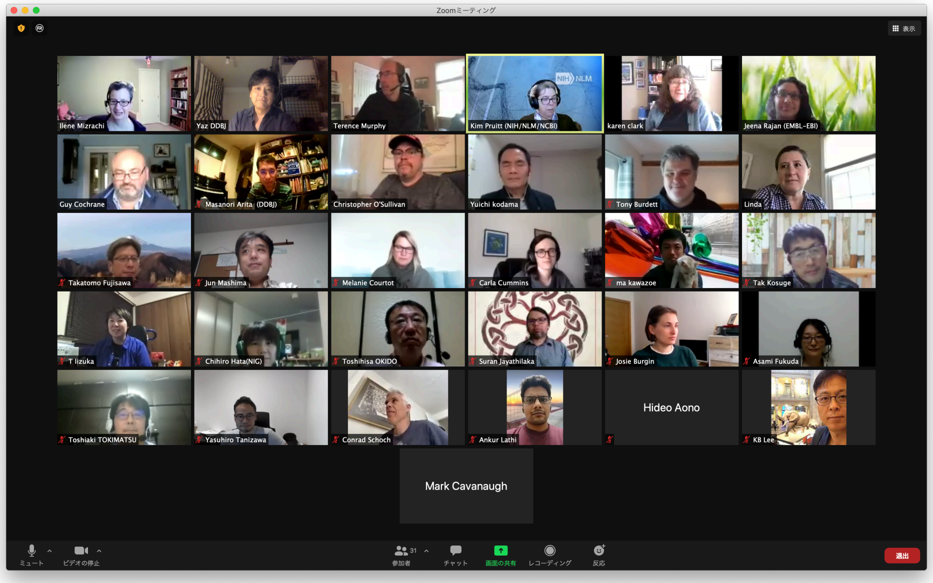
1995	EMBL data library was organized, and asked international cooperation for nucleotide sequence data bank to Japan.
1982	EMBL and GenBank started international cooperation, and invited Japan to participate their data bank.
1983	Aimed at contribution for international data bank to collect, to evaluate and to provide nucleotide sequence data, trial data loading was started.
1984	NIG; the National Institute of Genetics was reorganized as an Inter-University Research Institute. DDBJ began to work at NIG.
1986	DNA Database Advisory Committee organized.
1987	DDBJ release 1 was provided. By this release, we regard this year as official start of DDBJ operation.
1995.04	To operate DDBJ more efficiently, CIB; the Center for Information Biology was established in NIG.
2001.04	CIB was reorganized as CIB-DDBJ; the Center for Information Biology and DNA Data Bank of Japan
2004.04	NIG was reorganized as a member of ROIS; Research Organization of Information and Systems. DDBJ has also belonged to ROIS.
2005.05	DDBJ, EMBL-Bank and GenBank agreed to call their collaboration INSDC; International Nucleotide Sequence Database Collaboration; and to call the unified nucleotide sequence database INSD; the International Nucleotide Sequence Database.
2007.04	DBCLS; Database Center for Life Science was newly founded in ROIS
2009	DDBJ faculty staff have greatly been reshuffled. DDBJ collaborates with DBCLS more closely. INSDC added a collaborative meeting to deal with huge sequence data produced by the next generation sequencers (Sequence Read Archive) and traces produced by traditional sequencers (Trace Archive).
2012.04	DDBJ, expanding its DNA databank activities, was restructured as one of the Intellectual Infrastructure Project Centers of NIG, being separated from CIB.
2013.10	Collaborating with NBDC; National Bioscience Database Center, DDBJ Center started to operate the archive for all types of individual-level genetic and de-identified phenotypic data from human subjects, JGA; Japanese Genotype-phenotype Archive.

DDBJ (from Release note 115) 47
Jun Mashima, Kazunori Aoki, Hideo Aono, Yuji Ashizawa, Yukino Dobashi, Mayumi Ejima, Masahiro Fujimoto, Asami Fukuda, Tomohiro Hirai, Michiaki Hiramatsu, Naofumi Ishikawa, Kenji Kato, Aimi Kawasaki, Yuichi Kodama, Junko Kohira, Takehide Kosuge, Kyungbum Lee, Mika Maki, Fujitaka Matsumori, Kimiko Mimura, Hideki Mochizuki, Naoko Murakata, Yoshiyuki Nogi, Toshihisa Okido, Yoshihiro Okuda, Maki Ono, Katsunaga Sakai, Yukie Sakon, Makoto Sato, Rie Sugita, Kimiko Suzuki, Takahiro Suzuki, Daisuke Takagi, Yaeko Takiguchi, Toshiaki Tokimatsu, Haru Tsutsui, Koji Watanabe, Tomohiko Yasuda, Emi Yokoyama, Masanori Arita, Takeshi Kawashima, Osamu Ogasawara, Kosaku Okubo, Nozomu Sakurai, Yasuhiro Tanizawa, Toshihisa Takagi, and Yasukazu Nakamura

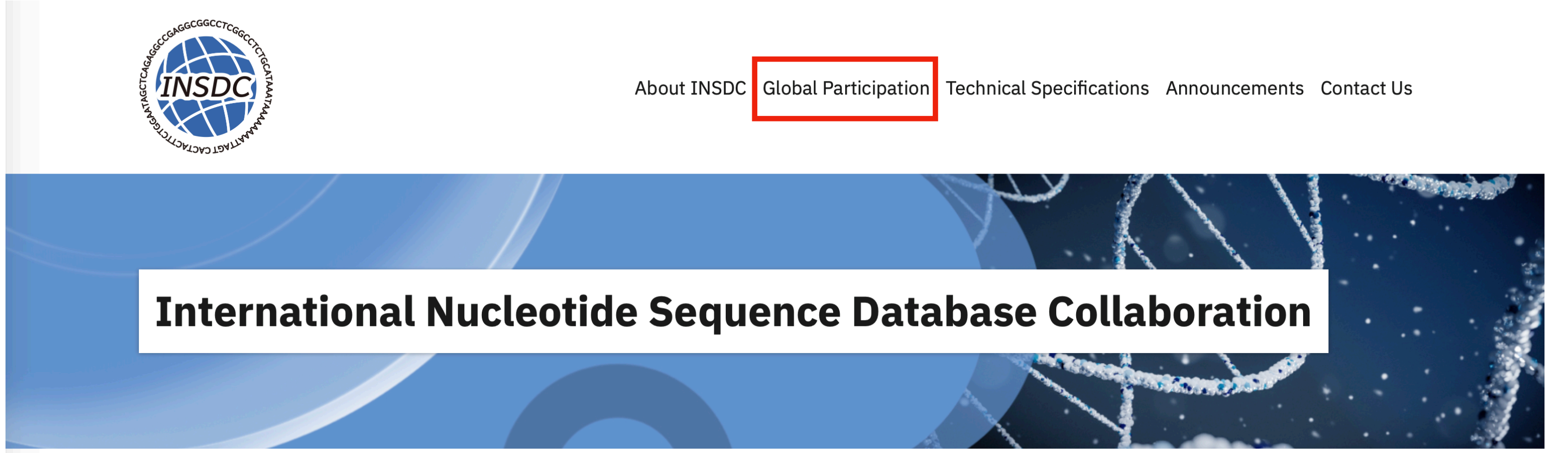
ENA (from Release note 115) 27
Blaise Alako, Clara Amid, Lawrence Bower, Ana Cerdeno-Taraga, Iain Cleland, Richard Gibson, Neil Goodgame, Petra ten Hoopen, Mikyung Jang, Simon Kay, Rasko Leinonen, Xin Liu, Arnaud Oisel, Rodrigo Lopez, Hamish McWilliam, Nima Pakseresht, Sheila Plaister, Rajesh Radhakrishnan, Kethy Reddy, Stephane Riviere, Marc Rossello, Nicole Silvester, Dmitriy Smirnov, Ana Luisa Toribio, Daniel Vaughan, Vadim Zalunin and Guy Cochrane

GenBank (from Release note 227) 80
Mark Cavanaugh, Ilene Mizrachi, Michael Baxter, Shelby Bidwell, Lori Black, Larissa Brown, Vincent Calhoun, Larry Chlumsky, Karen Clark, Jianli Dai, Scott Durkin, Francescopaolo di Cello, Michel Eschenbrenner, Michael Fetchko, Linda Frisse, Andrea Gocke, Anjanette Johnston, Mark Landree, Jason Lowry, Richard McVeigh, DeAnne Olsen Cravaritis, Leigh Riley, Susan Schafer, Beverly Underwood, and Linda Yankie, Serge Bazhin, Evgueni Belyi, Colleen Bollin, Yoon Choi, Sergey Dikunov, Ilya Dondoshansky, Justin Foley, Viatcheslav Gorelenkov, Sergiy Gotvyanskyy, Eugene Gribov, Jonathan Kans, Leonid Khotomliansky, Michael Kimelman, Dmitri Kishchukov, Michael Kornbluh, Alex Kotliarov, Alexey Kuznetsov, Frank Ludwig, Anatoly Mnev, Jim Ostell, Vasuki Palanigobu, Anton Perkov, Andriy Petrow, Sergey Petrunin, Wenyao Shi, Denis Sinyakov, Thomas Smith, Vladimir Soussov, Elena Starchenko, Hanzhen Sun, Andrei Vereshchagin, Jewen Xiao, Eugene Yaschenko, Liwei Zhou, Slava Khotomliansky, Igor Lozitskiy, Craig Oakley, Eugene Semenov, Ben Slade, Constantin Vasilyev, Sherri Bailey, William Bocik, David Brodsky, Peter Cooper, Jada Lewis, Hanguan Liu, Bonnie Maidak, Wayne Matten, Scott McGinnis, Rana Morris, Monica Romiti, Eric Sayers, Tao Tao, Majda Valjavec-Gratian and Kim Pruitt

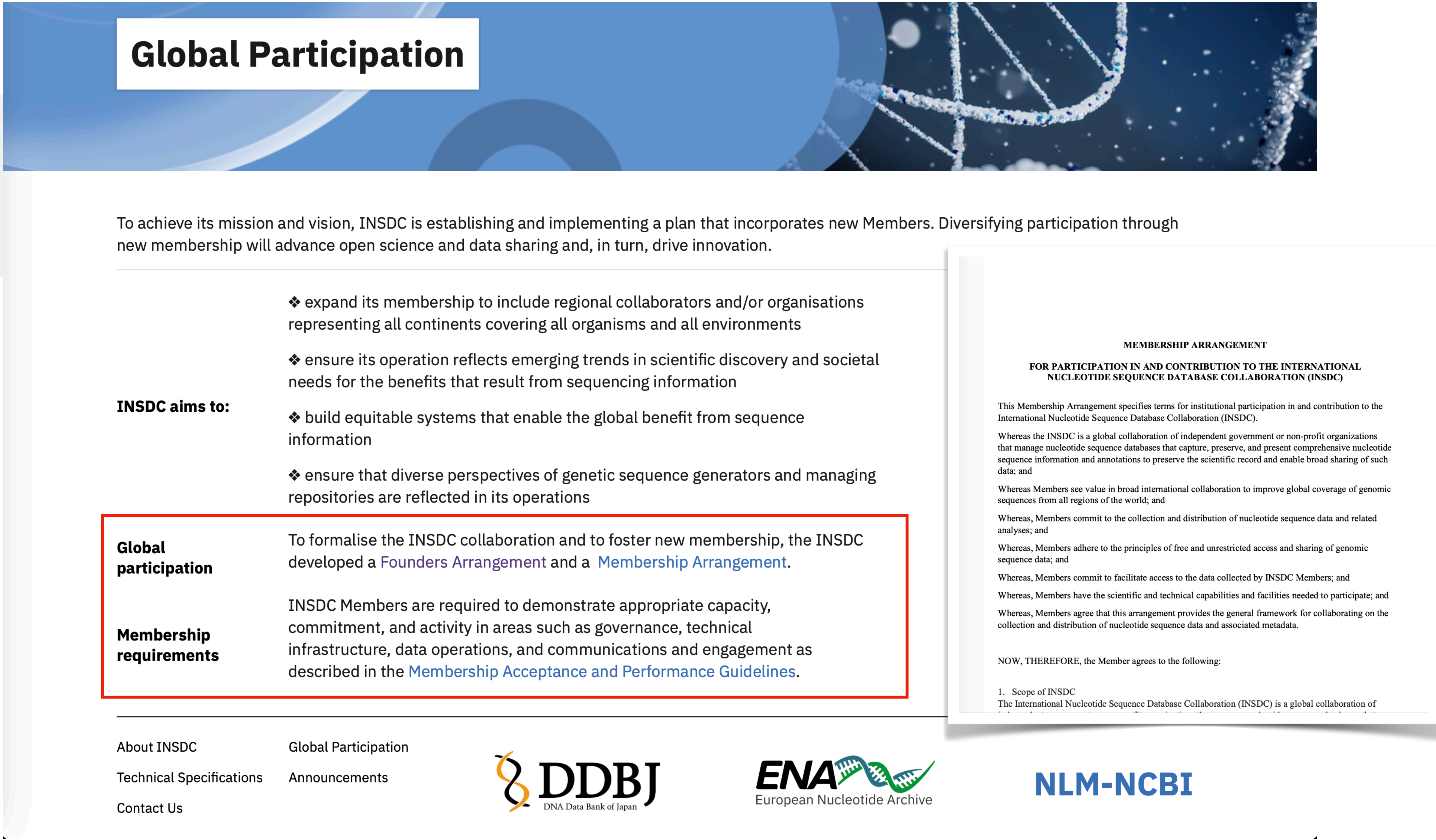
The INSDC meeting at EBI, May 2019



The INSDC meeting online, May 2021



International Nucleotide Sequence Database Collaboration



Global Participation

To achieve its mission and vision, INSDC is establishing and implementing a plan that incorporates new Members. Diversifying participation through new membership will advance open science and data sharing and, in turn, drive innovation.

INSDC aims to:	❖ expand its membership to include regional collaborators and/or organisations representing all continents covering all organisms and all environments ❖ ensure its operation reflects emerging trends in scientific discovery and societal needs for the benefits that result from sequencing information ❖ build equitable systems that enable the global benefit from sequence information ❖ ensure that diverse perspectives of genetic sequence generators and managing repositories are reflected in its operations
Global participation	To formalise the INSDC collaboration and to foster new membership, the INSDC developed a Founders Arrangement and a Membership Arrangement.
Membership requirements	INSDC Members are required to demonstrate appropriate capacity, commitment, and activity in areas such as governance, technical infrastructure, data operations, and communications and engagement as described in the Membership Acceptance and Performance Guidelines.

MEMBERSHIP ARRANGEMENT

FOR PARTICIPATION IN AND CONTRIBUTION TO THE INTERNATIONAL NUCLEOTIDE SEQUENCE DATABASE COLLABORATION (INSDC)

This Membership Arrangement specifies terms for institutional participation in and contribution to the International Nucleotide Sequence Database Collaboration (INSDC).

Whereas the INSDC is a global collaboration of independent government or non-profit organizations that manage nucleotide sequence databases that capture, preserve, and present comprehensive nucleotide sequence information and annotations to preserve the scientific record and enable broad sharing of such data; and

Whereas Members see value in broad international collaboration to improve global coverage of genomic sequences from all regions of the world; and

Whereas, Members commit to the collection and distribution of nucleotide sequence data and related analyses; and

Whereas, Members adhere to the principles of free and unrestricted access and sharing of genomic sequence data; and

Whereas, Members commit to facilitate access to the data collected by INSDC Members; and

Whereas, Members have the scientific and technical capabilities and facilities needed to participate; and

Whereas, Members agree that this arrangement provides the general framework for collaborating on the collection and distribution of nucleotide sequence data and associated metadata.

NOW, THEREFORE, the Member agrees to the following:

1. Scope of INSDC

The International Nucleotide Sequence Database Collaboration (INSDC) is a global collaboration of

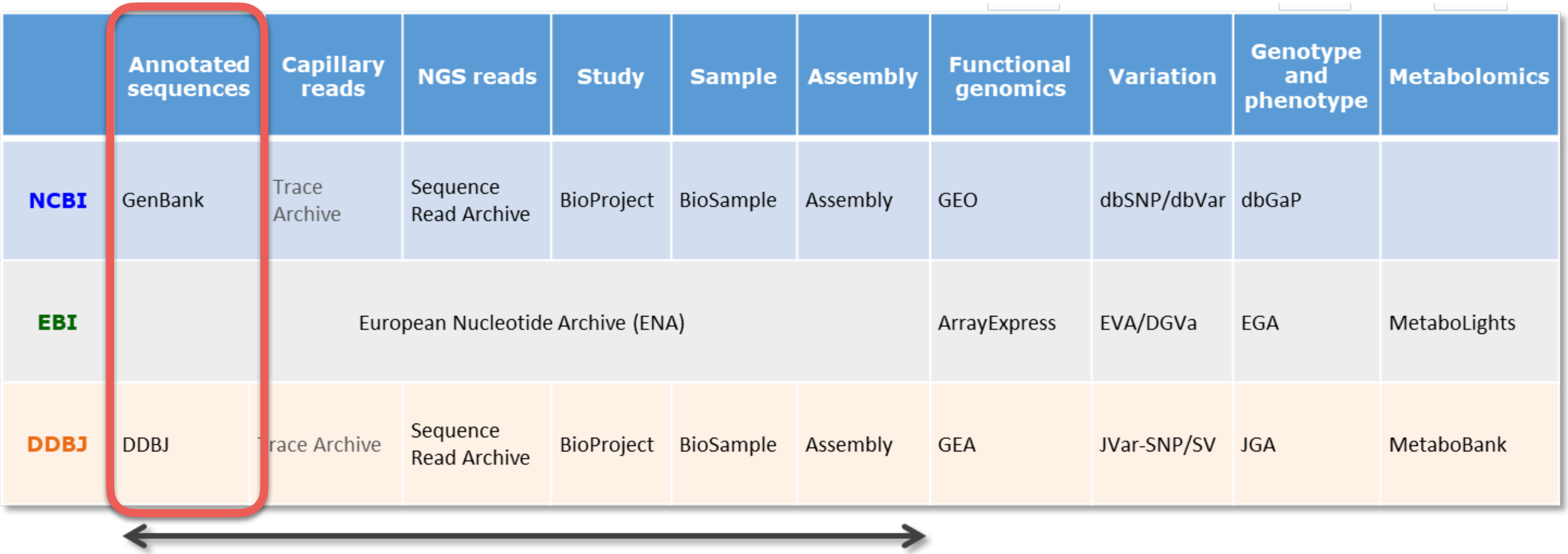
Activities on the databases

Databases at Bioinformation and DDBJ Center

	Annotated sequences	Capillary reads	NGS reads	Study	Sample	Assembly	Functional genomics	Variation	Genotype and phenotype	Metabolomics
NCBI	GenBank	Trace Archive	Sequence Read Archive	BioProject	BioSample	Assembly	GEO	dbSNP/dbVar	dbGaP	
EBI	European Nucleotide Archive (ENA)						ArrayExpress	EVA/DGVa	EGA	MetaboLights
DDBJ	DDBJ	Trace Archive	Sequence Read Archive	BioProject	BioSample	Assembly	GEA	JVar-SNP/SV	JGA	MetaboBank

INSDC：International Nucleotide Sequence Databank Collaboration

Databases at Bioinformation and DDBJ Center



	Annotated sequences	Capillary reads	NGS reads	Study	Sample	Assembly	Functional genomics	Variation	Genotype and phenotype	Metabolomics
NCBI	GenBank	Trace Archive	Sequence Read Archive	BioProject	BioSample	Assembly	GEO	dbSNP/dbVar	dbGaP	
EBI	European Nucleotide Archive (ENA)						ArrayExpress	EVA/DGVa	EGA	MetaboLights
DDBJ	DDBJ	Trace Archive	Sequence Read Archive	BioProject	BioSample	Assembly	GEA	JVar-SNP/SV	JGA	MetaboBank

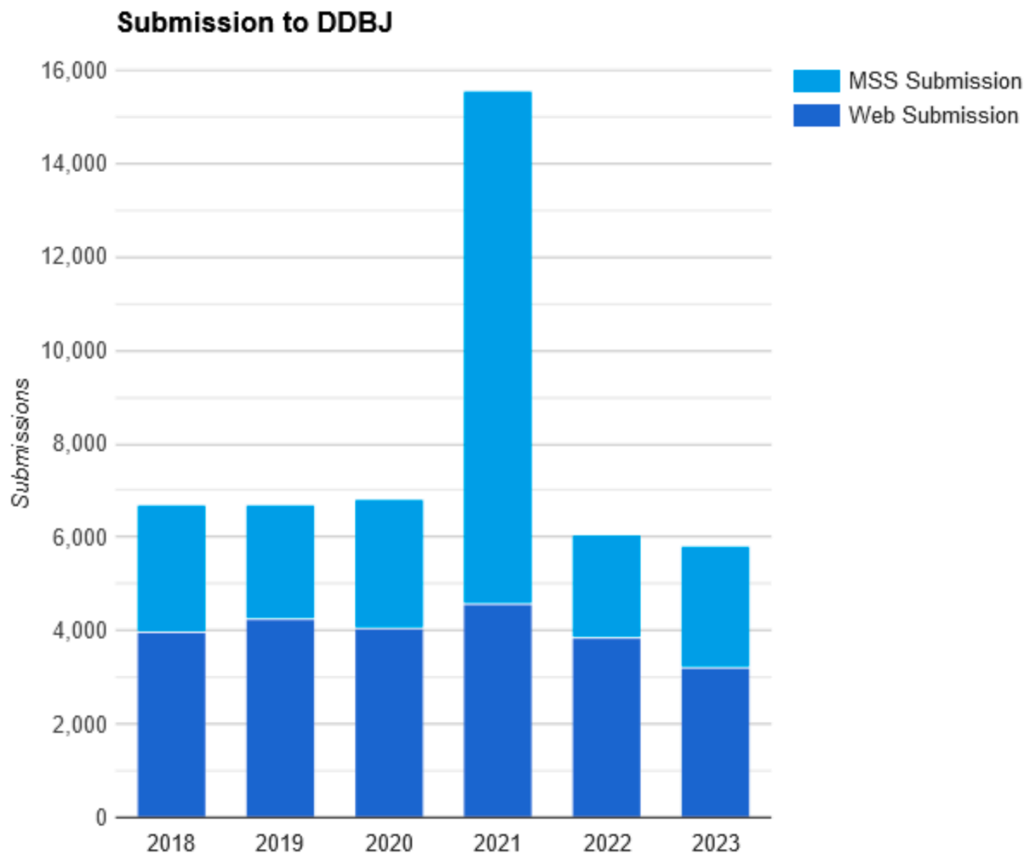
INSDC : International Nucleotide Sequence Databank Collaboration

An example for the “trad” database

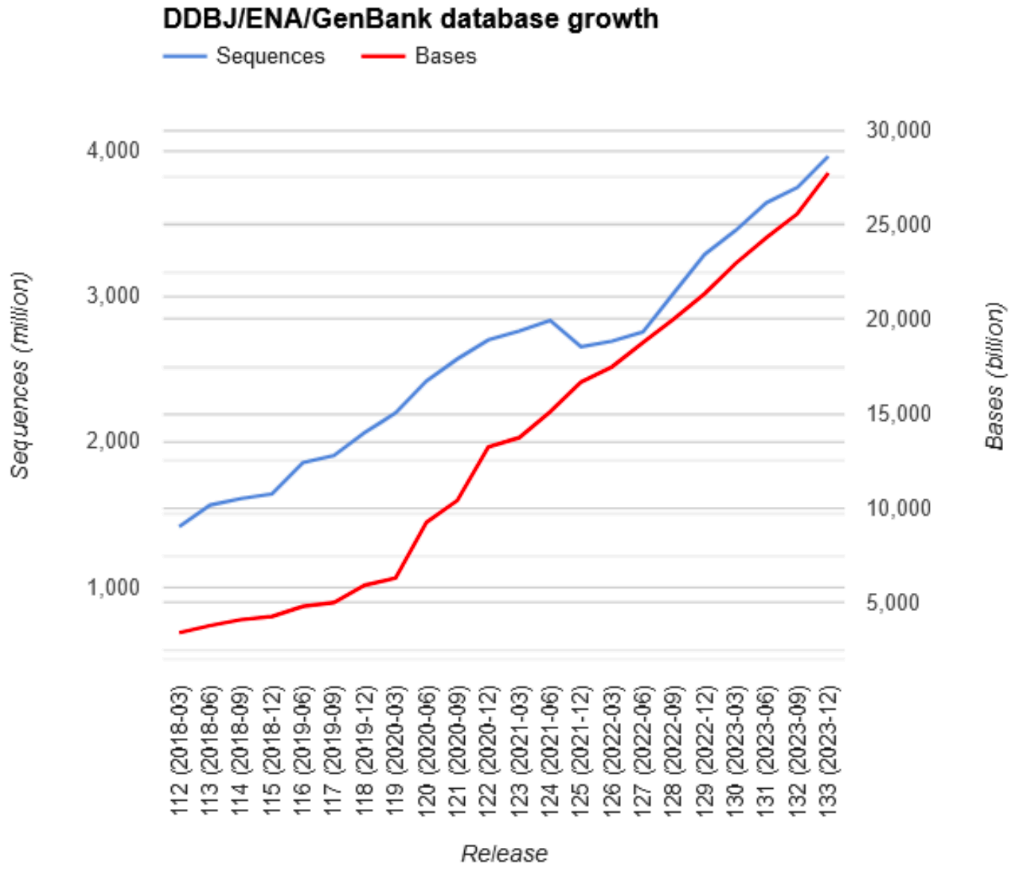
```
LOCUS       AB091058                2109 bp    DNA     linear   BCT 02-SEP-2003
DEFINITION  Gluconacetobacter xylinus cmcase, ccp genes for
            endo-beta-1,4-glucanase, cellulose complementing protein, complete
            cds.
ACCESSION   AB091058
VERSION     AB091058.1
KEYWORDS    .
SOURCE      Gluconacetobacter xylinus
  ORGANISM  Gluconacetobacter xylinus
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
            Acetobacteraceae; Gluconacetobacter.
REFERENCE   1  (bases 1 to 2109)
  AUTHORS   Kawano,S., Tajima,K., Uemori,Y., Yamashita,H., Erata,T.,
            Munekata,M. and Takai,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-2002) to the DDBJ/EMBL/GenBank databases.
            Contact:Kenji Tajima
            Hokkaido University, Graduate School of Engineering; N13W8,
            Kita-ku, Sapporo, Hokkaido 060-8628, Japan
REFERENCE   2
  AUTHORS   Kawano,S., Tajima,K., Uemori,Y., Yamashita,H., Erata,T.,
            Munekata,M. and Takai,M.
  TITLE     Cloning of Cellulose Synthesis Related Genes from Acetobacter
            xylinum ATCC23769 and ATCC53582: Comparison of Cellulose Synthetic
            Ability Between ATCC23769 and ATCC53582
  JOURNAL   Unpublished (2002)
COMMENT
FEATURES             Location/Qualifiers
     source          1..2109
                     /db_xref="taxon:28448"
                     /mol_type="genomic DNA"
                     /note="synonym:Acetobacter xylinum"
                     /organism="Gluconacetobacter xylinus"
                     /strain="ATCC 53582"
     CDS             10..1038
                     /codon_start=1
                     /gene="cmcase"
                     /product="endo-beta-1,4-glucanase"
                     /protein_id="BAC82540.1"
                     /transl_table=11
                     /translation="MSVMAAMGGAQVLSSTGAFADTAPDAVAQQWAIFRAKYLRPSGR
                     VVDTGNGGESHSEGQGYGMLFAASAGDLASFQSMWMWARTNLQHTNDKLFSWRFLKGH
                     QPPVPDKNNATDGDLLIALALGRAGKRFQRPDYIQDAMAIYGDVLNLMTMKAGPYVVL
                     MPGAVGFTKKDSVILNLSYYVMPSLLQAFDLTADPRWRQVMEDGIRLVSAGRFGQWRL
                     PPDWLAVNRATGALSIASGWPPRFSYDAIRVPLYFYWAHMLAPNVLADFTRFWNNFGA
                     NALPGWVDLTTGARSPYNAPPGYLAVAECTGLDSAGELPTLDHAPDYYSAALTLLVYI
                     ARAEETIK"
     CDS             1035..2096
                     /codon_start=1
                     /gene="ccp"
                     /product="cellulose complementing protein"
                     /protein_id="BAC82541.1"
                     /transl_table=11
                     /translation="MSASGSDEVAGGGQAGSPQDFQRVLRSFGVEGGQYSYRPFVDRS
                     FDVTGVPEAVERHFDQAEHDTAVEEQVTPAPQIAVAPPPPPVVPDPPAIVTETAPPPP
                     VVVSAPVTYEPPAAAVPAEPPVQEAPVQAAPVPPAPVPPIAEQAPPAAPDPASVPYAN
                     VAAAPVPPDPAPVTPAPQARVTGPNTRMVEPFSRPQVRTVQEGATPSRVPSRSMNAFP
                     RTSASSISERPVDRGVADEWSPVPKARLSPRERPRPGDLSFFFQGMRDTRDEKKFFPV
                     ASTRSVRSNVSRMTSMTKTDTNSSQASRPGSPVASPDGSPTMAEVFMTLGGRATELLS
                     PRPSLREALLRRRENEEES"
BASE COUNT          343 a          661 c          661 g          444 t
ORIGIN
        1 cgttccttta tgtcggtcat ggcggcgatg ggaggggcgc aggtgctttc atccaccggt
       61 gcgttcgcag acaccgcccc cgatgcggtc gcgcagcaat gggccatctt ccgcgccaag
      121 tatcttcgtc ccagcggacg tgtcgtggat acgggcaatg gtggcgaatc ccatagtgag
      181 gggcagggct atggcatgct ctttgccgcg tcggcggggg accttgcgtc gttccagtcg
      241 atgtggatgt gggcgcgcac caacctgcag cataccaatg acaagctgtt ttcctggcgg
      301 ttcctcaagg ggcatcagcc cccggtgccc gacaagaaca atgccacaga tggcgacctg
      361 ctgatcgcgc ttgcgcttgg tcgtgcgggc aagcgtttcc agcgccccga ttacattcag
      421 gacgccatgg ccatttatgg cgatgtgctg aacctgatga cgatgaaggc gggaccgtat
      481 gtcgtcctca tgcccggtgc tgtcggcttt accaagaagg acagcgtgat cctcaacctg
      541 tcctattacg tcatgccctc gctgctgcag gcgttcgacc ttacggccga cccgcgctgg
      601 cgtcaggtga tggaagacgg gattcgcctt gtttccgccg gccgtttcgg gcagtggcgc
      661 ctgccccccg actggctggc ggtgaatcgc gccaccggtg cgctgtcgat cgcatcggga
      721 tggccgccgc gcttttccta tgatgcgatt cgggtgccgc tttattttta ttgggcgcat
      781 atgctggcgc cgaacgtgtt ggctgatttc acccgattct ggaataattt cggggctaat
      841 gccctgccag gatgggttga tctgacaaca ggggcgcgtt cgccgtacaa cgccccgcct
      901 ggatatcttg ctgttgccga atgcacgggg cttgattctg ccggggaact cccgacactg
      961 gatcatgcgc ccgattatta ttccgcagcg ttgacgctgc tcgtttacat cgcgcgggcg
     1021 gaggagacta taaagtgagt gcttcagggt ctgatgaggt ggctgggga gggcaggctg
     1081 gaagtccgca ggattttcag cgggtcctgc gttcttttgg tgtcgaaggt gggcagtatt
     1141 cctaccggcc gtttgttgac cgttcctttg atgtgacagg cgtgcccgag gctgttgaaa
     1201 ggcacttcga tcaggcggag catgacacgg cggttgagga gcaggtcact cccgcgccac
     1261 aaatcgcggt cgcaccgcca ccgccgccag tcgttcctga cccgcccgcc atcgtgacgg
     1321 aaaccgcgcc cccgccgcct gtcgtggtca gcgctccggt cacgtatgaa cccccggctg
     1381 ccgccgtgcc ggcagagcct cccgttcagg aagcccccgt gcaggcggcg ccggttcccc
     1441 ccgcgcctgt gcccccgatt gcggagcagg ctcctcccgc ggcgccggac ccggcatccg
     1501 tgccgtatgc gaacgtcgcg gcagcacccg ttccacctga tcccgcaccg gttacgcctg
     1561 cgccgcaggc gcgcgtgacg gggccgaaca cccgtatggt ggagcccttt tcccgcccgc
     1621 aggtccgcac ggtgcaggag ggggcaaccc cgtcacgtgt accttcgcgt tcaatgaacg
     1681 ctttcccccg cacatcagca tcgtccataa gtgagcgtcc ggtggacagg ggtgttgccg
     1741 atgaatggag tcctgttccg aaggcacgcc tcagcccgcg ggagcgtccg cgtcccggcg
     1801 atctgagctt tttctttcag gggatgcgcg acacccgtga tgaaaagaag ttctttcccg
     1861 tggcgtccac gcgatcagtt cgttctaatg tttccaggat gaccagcatg accaagacag
     1921 acacgaattc ctctcaggct tctcgtcccg gcagccccgt cgcctcgcct gatgggtcgc
     1981 ccacaatggc cgaagtgttc atgacgctgg gtggtcgtgc gacggaactc ctcagccccc
     2041 gtccttcgct gcgggaggcg ctgttgcgtc gtcgtgaaaa cgaagaagaa tcctaaggcc
     2101 ctatattca

//
```

Trad in DDBJ



Annual number of registrations to Trad DDBJ.



Total number of sequences and bases released by DDBJ.

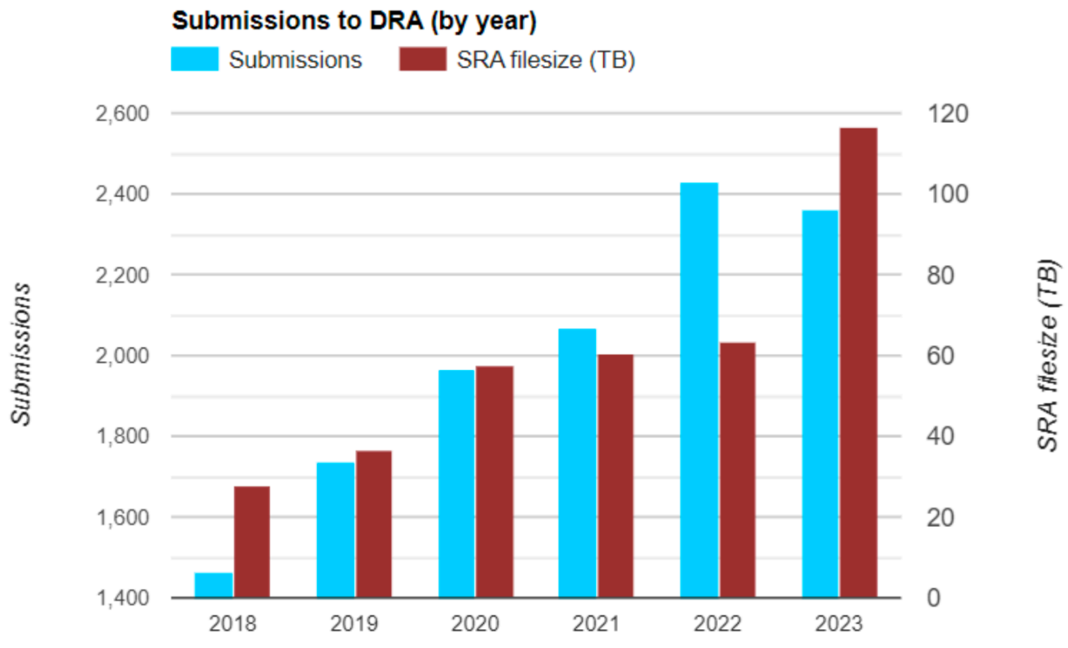
- The total number of registrations related to DDBJ entries in 2023 were ca. 6,000; about a half of submissions made by web interface and a half were pre-format MSS submittions. (The rapid increase in the number of MSS submissions in 2021 was due to a large amount of MAG (Metagenome-Assembled Genome) from the University of Tokyo.) As of December 2023, the total number of bases is about 27 trillion, and the total number of sequences is about 4.0 billion.

Databases at Bioinformation and DDBJ Center

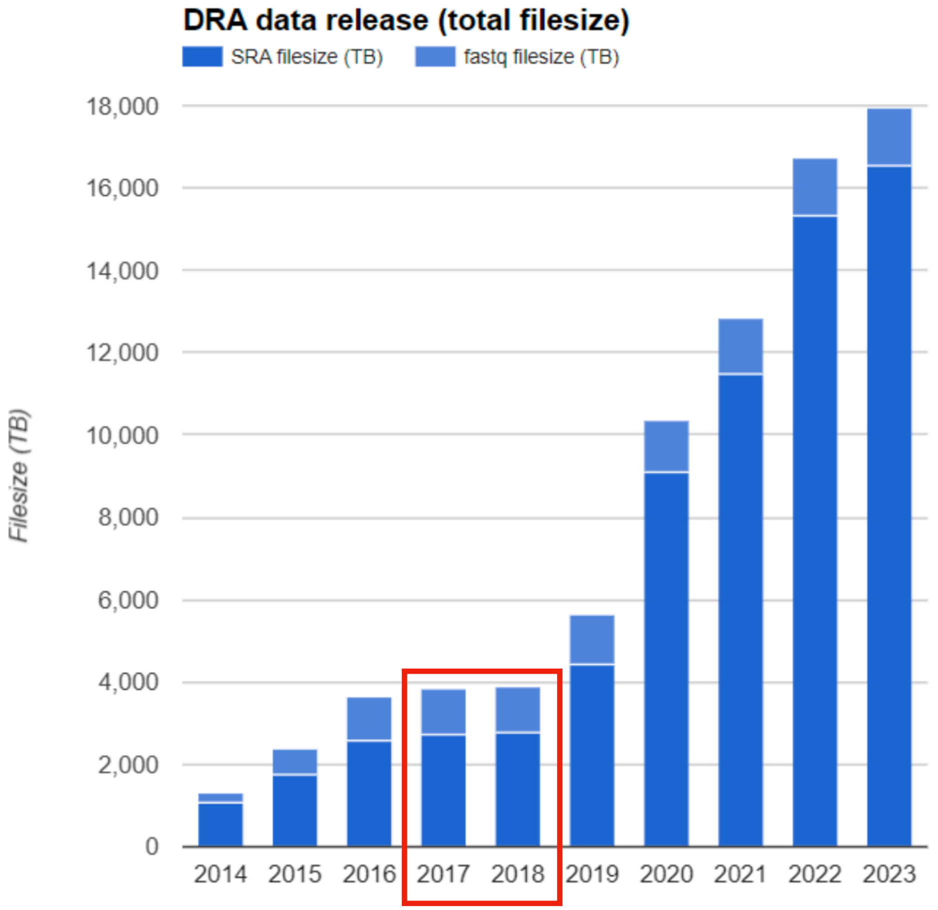
	Annotated sequences	**Capillary reads**	**NGS reads**	**Study**	**Sample**	**Assembly**	**Functional genomics**	**Variation**	**Genotype and phenotype**	**Metabolomics**
NCBI	GenBank	Trace Archive	Sequence Read Archive	BioProject	BioSample	Assembly	GEO	dbSNP/dbVar	dbGaP	
EBI	European Nucleotide Archive (ENA)						ArrayExpress	EVA/DGVa	EGA	MetaboLights
DDBJ	DDBJ	Trace Archive	Sequence Read Archive	BioProject	BioSample	Assembly	GEA	JVar-SNP/SV	JGA	MetaboBank

INSDC: International Nucleotide Sequence Databank Collaboration

SRA in DDBJ (DRA)



Number of registrations and amount of files per year



Amount of SRA and fastq files in DRA

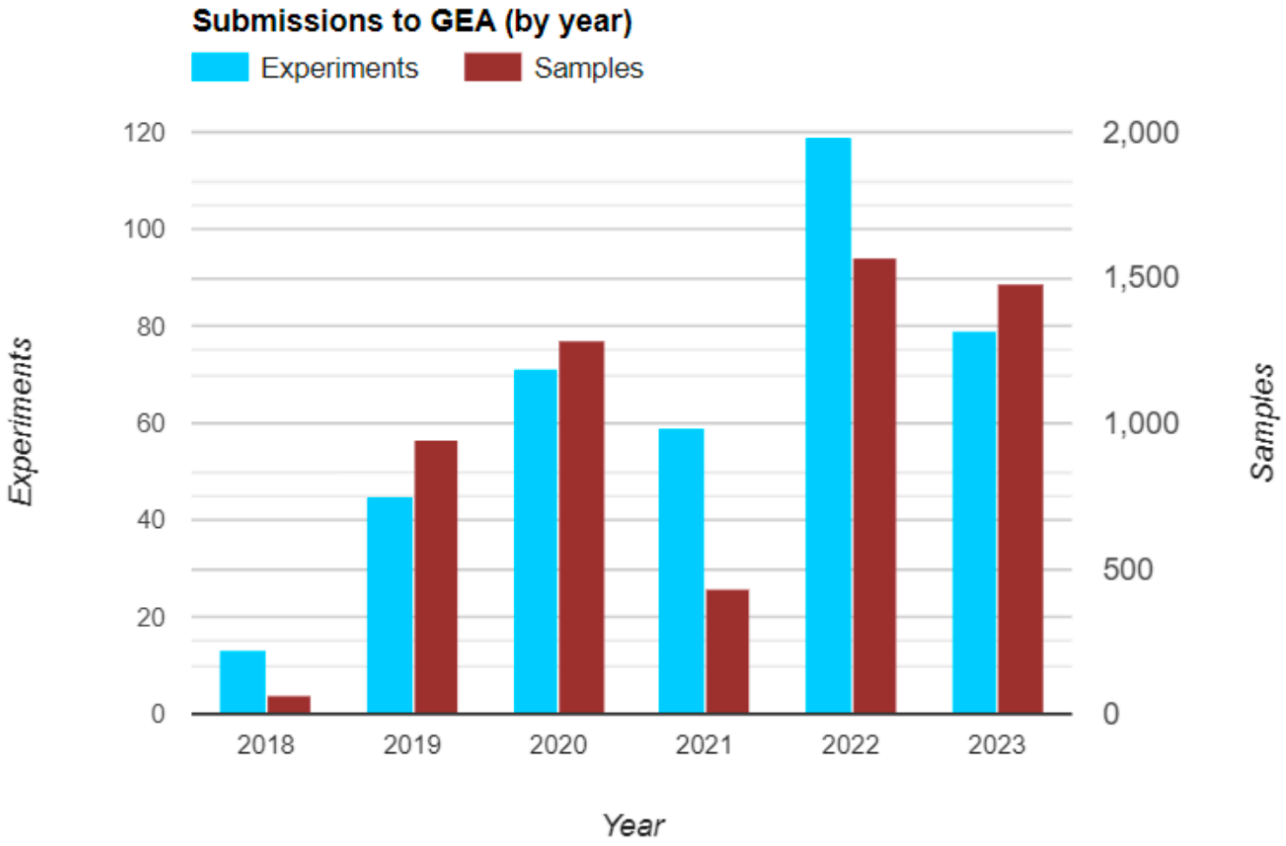
- DDBJ SRA (DRA) registrations in 2023 were 2,361 (116TB). The total SRA and fastq file sizes published were 16.5PB and 1.4PB, respectively. (From April 2017 to May 2019, the amount did not grow due to the cessation of mirroring of EBI/NCBI data.)

Databases at Bioinformation and DDBJ Center

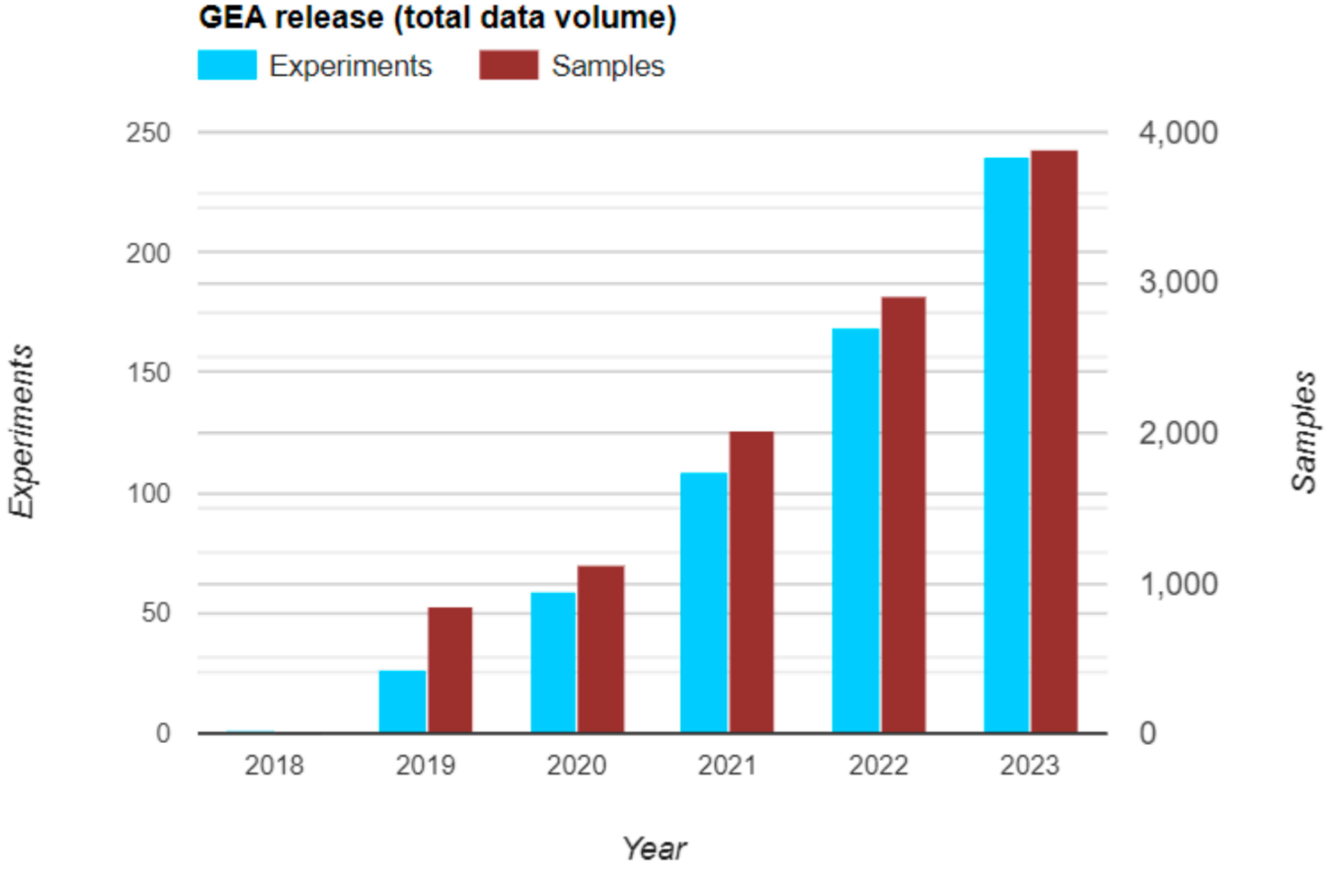
	Annotated sequences	Capillary reads	NGS reads	Study	Sample	Assembly	Functional genomics	Variation	Genotype and phenotype	Metabolomics
NCBI	GenBank	Trace Archive	Sequence Read Archive	BioProject	BioSample	Assembly	GEO	dbSNP/dbVar	dbGaP	
EBI	European Nucleotide Archive (ENA)						ArrayExpress	EVA/DGVa	EGA	MetaboLights
DDBJ	DDBJ	Trace Archive	Sequence Read Archive	BioProject	BioSample	Assembly	GEA	JVar-SNP/SV	JGA	MetaboBank

INSDC：International Nucleotide Sequence Databank Collaboration

GEA: Genomic Expression Archive



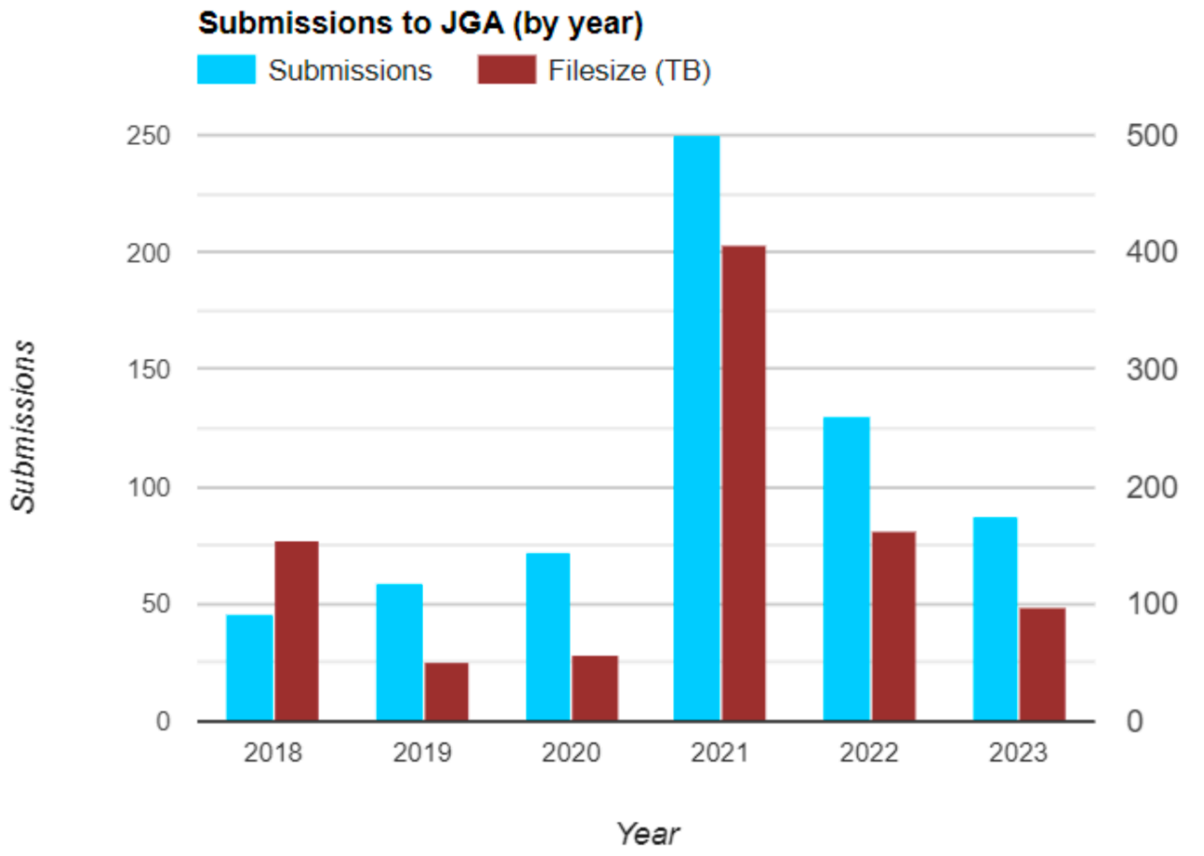
GEA's submissions and samples numbers by year.



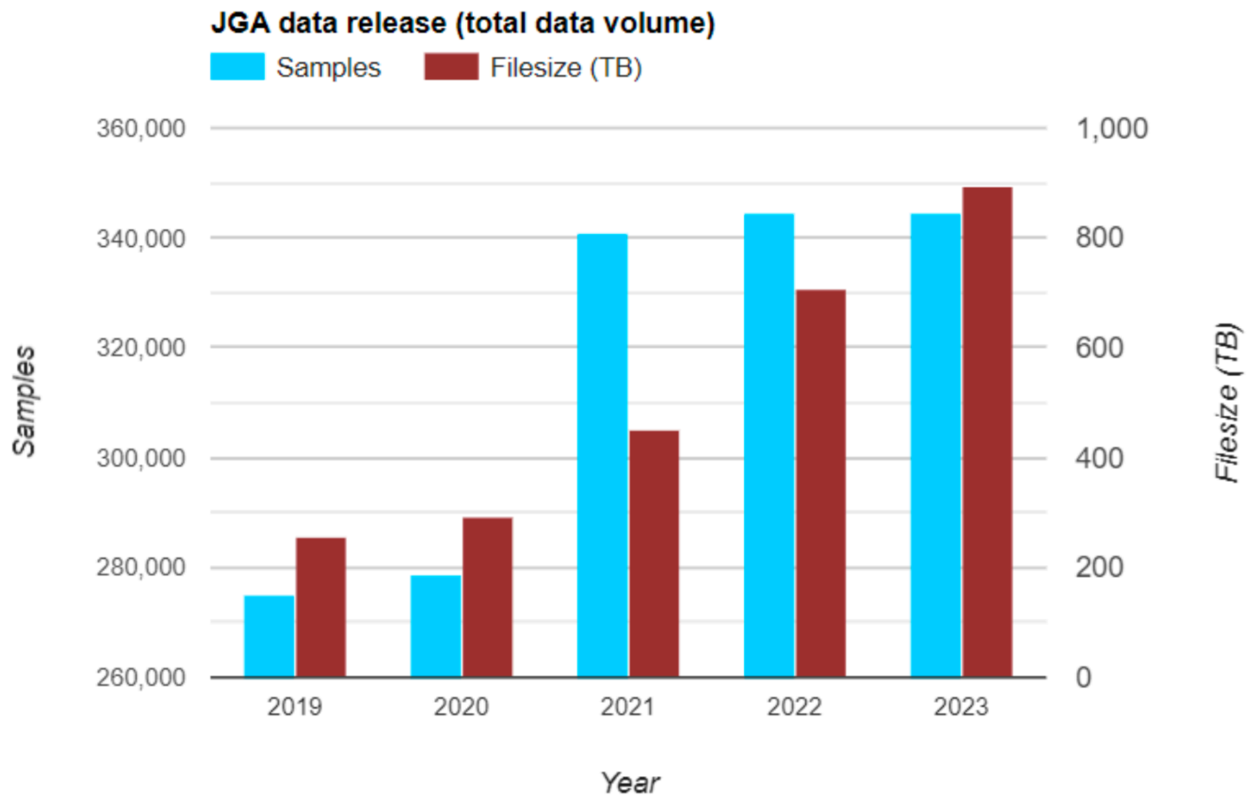
Number of relased experiments and samples from GEA.

- GEA is a database of gene expression data from microarrays and NGS in DDBJ. In 2023, we had 79 data submissions, and 240 Experiments and ca. 4,000 Samples in total, were publicly available.

Japanese Genotype-phenotype Archive (JGA)



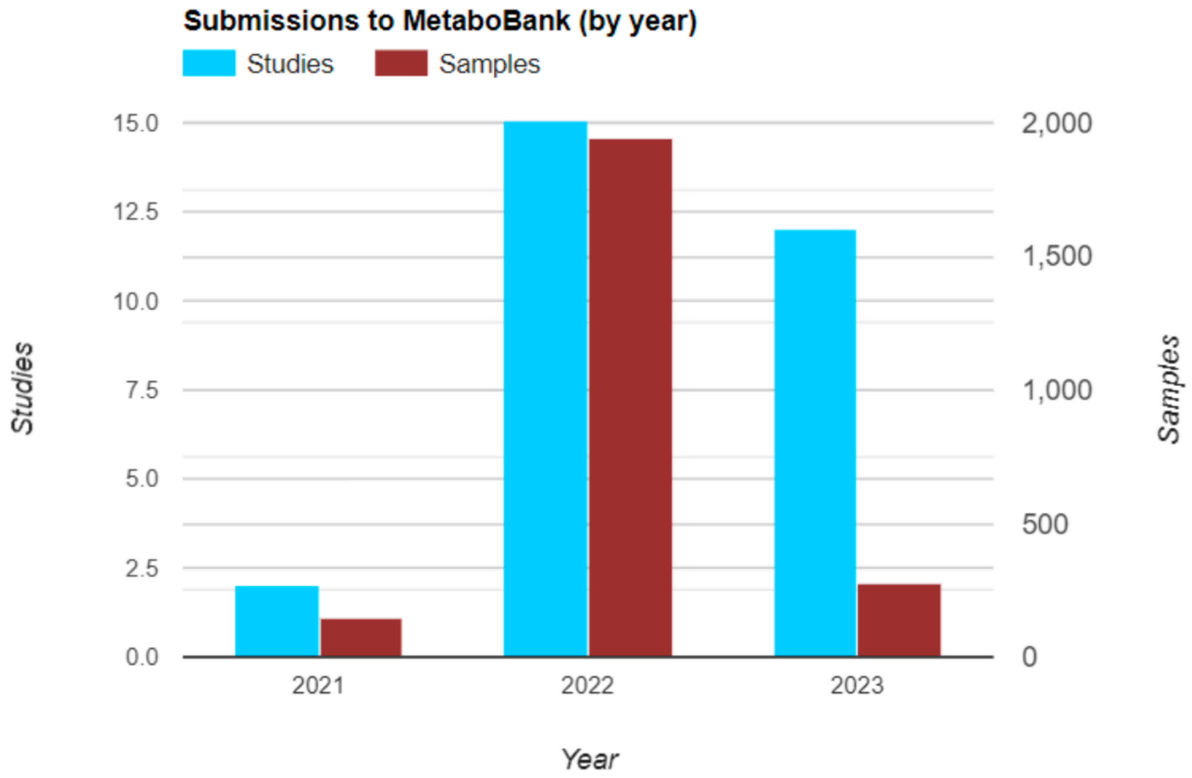
JGA's submissions and samples numbers by year.



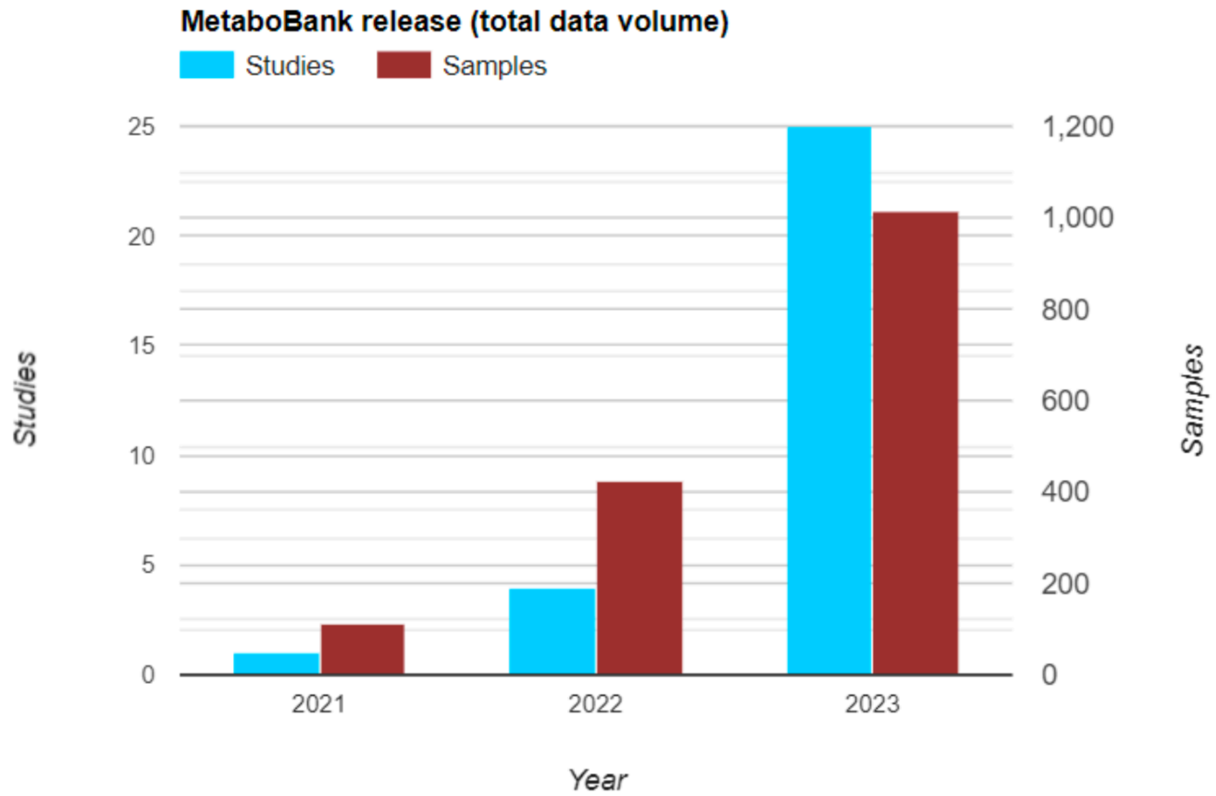
Number of relased experiments and samples from JGA.

- A control-access database for personal genome phenotype/genotype information operated jointly with NBDC. Registration and use of the database are subject to NBDC review. In 2023, 87 data and 97 TB were submitted. As of the end of 2023, 358 studies, 344,702 samples, and 895 TB of data were released. The number of applications approved by NBDC in 2022 totaled 49 (domestic: 17, overseas: 18). A cumulative total number is 263 applications approved.

MetaboBank



MetaboBank submissions and sample numbers by year.

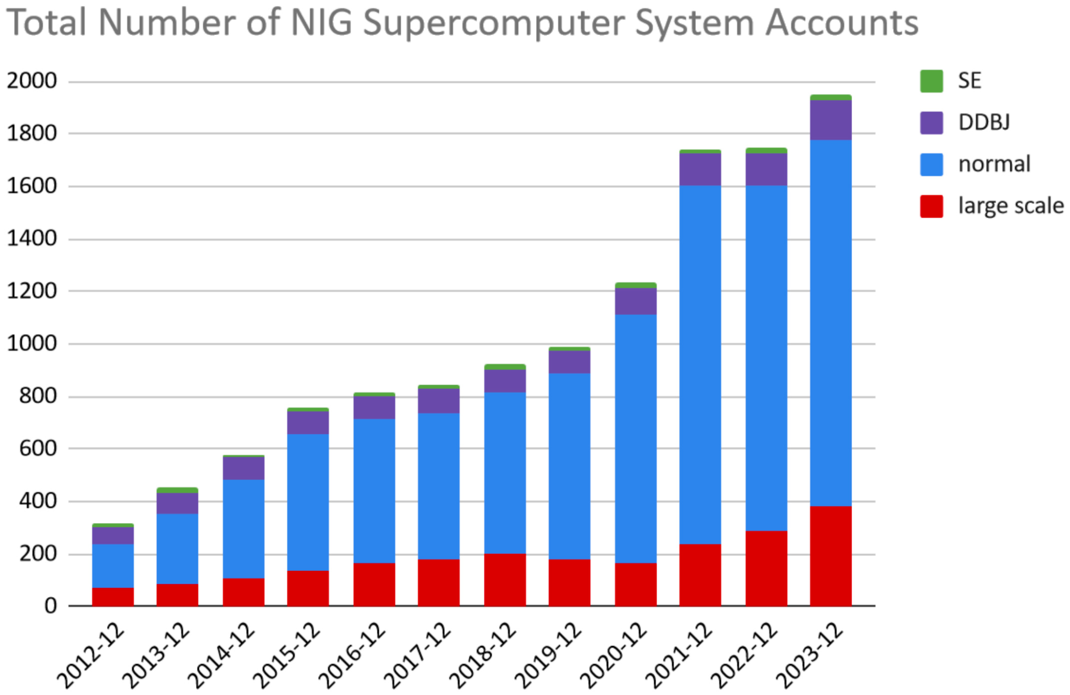


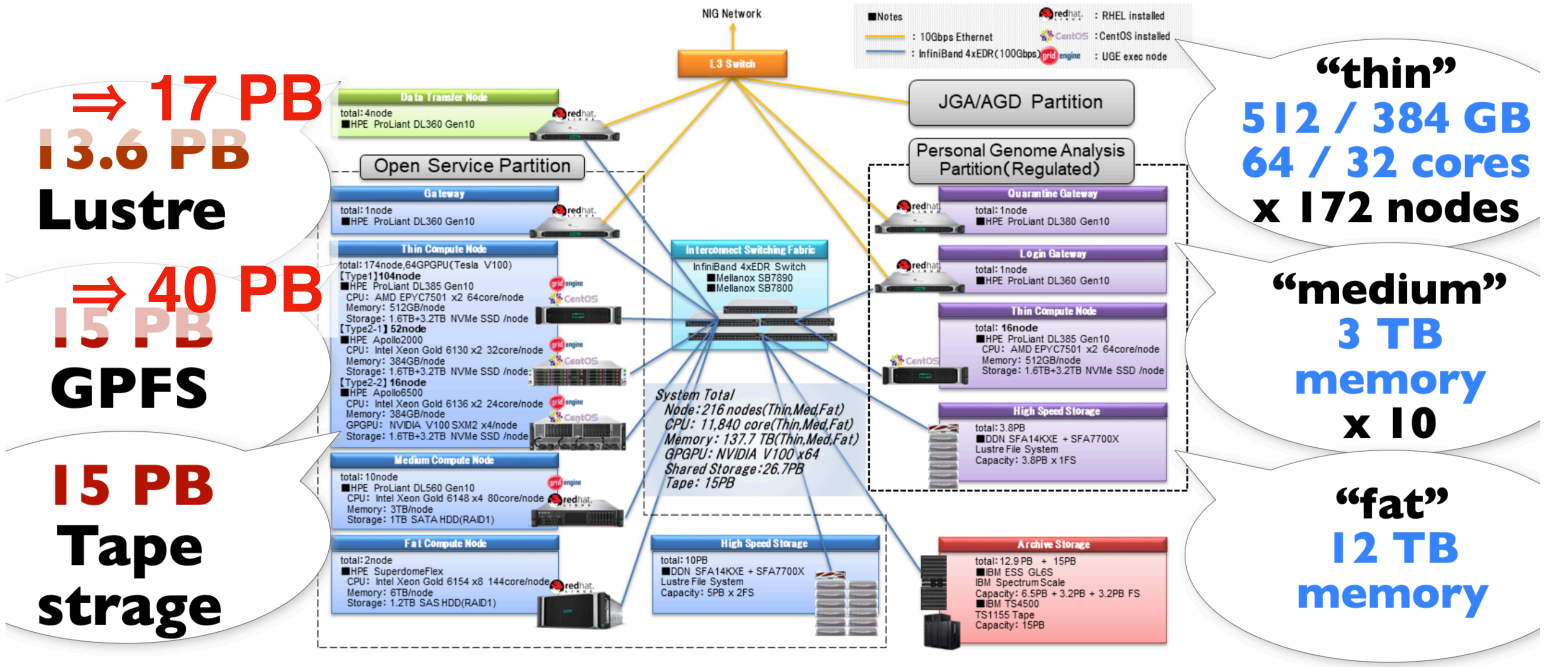
The number of released studies and samples from MetaboBank.

- MetaboBank Version 2 uses an IDF/SDRF registration form as a metadata input format equivalent to GEA and EBI's MetaboLight. Web content, including the registration form and guidelines and the workflow from data registration to publication, have been prepared and publicized. The data registration for 2023 is 12 Studies, and 25 Studies were open at the end of 2023.

NIG Supercomputer: an infrastructure for lifescience

Osamu Ogasawara, PhD
Project Associate Professor
Head of HPC Division





⇒ 17 PB
~~13.6 PB~~
Lustre

⇒ 40 PB
~~15 PB~~
GPFS

15 PB
Tape
strage

72 PB

“thin”
512 / 384 GB
64 / 32 cores
x 172 nodes

“medium”
3 TB
memory
x 10

“fat”
12 TB
memory

14,336 core

- Singularity/Docker containers are available for users.
- As of last year (December 31, 2023), there were 1,953 total users.
- The large-capacity archive storage system dedicated to the database was a hierarchical storage system with 15 PB of disk and 15 PB of tape. It was replaced with a Lustre file system with an effective capacity of 40 PB (DDN, Hitachi, Ltd.) As a result, it is possible to mount DRAs and other files from all computation nodes of the supercomputer at a higher speed. The 15 PB tape part was purchased and continues to be used as an offline backup.

Capacity Buildings: Vietnam, Indonesia and Thailand (2023)

Japan-Vietnam Workshop ▶ すべて再生

1.1) Data repository and DDBJ
DDBJvideo
620 回視聴・1 年前

1.2) UNIX commands and programming environment
DDBJvideo
349 回視聴・1 年前

1.3) What you can do with the super-computer in DDBJ
DDBJvideo
258 回視聴・1 年前

2.1) Introduction to Microbial Genome Analysis
DDBJvideo
1045 回視聴・1 年前

2.2) Genome Analysis of Eukaryotes ~ Gene structur...
DDBJvideo
130 回視聴・6 か月前

Japan-Indonesia Workshop ▶ すべて再生

1.1) Data repository and DDBJ
DDBJvideo
620 回視聴・1 年前

1.2) UNIX commands and programming environment
DDBJvideo
349 回視聴・1 年前

1.3) What you can do with the super-computer in DDBJ
DDBJvideo
258 回視聴・1 年前

2.1) Introduction to Microbial Genome Analysis
DDBJvideo
1045 回視聴・1 年前

2.2) Genome Analysis of Eukaryotes ~ Gene structur...
DDBJvideo
211 回視聴・1 年前

第34回 DDBJing 講習会 in 三島（2016.12.16-17） ▶ すべて再生

NIG SuperComputer の利用方法(2.1) ログインして解析を

NIG SuperComputer の利用方法(2.2) ログインして解析を

BioProject, BioSample の紹介

NIG SuperComputer の利用方法(1)PCに公開鍵を設定して

DRA(DDBJ Read Sequence Archive) の紹介

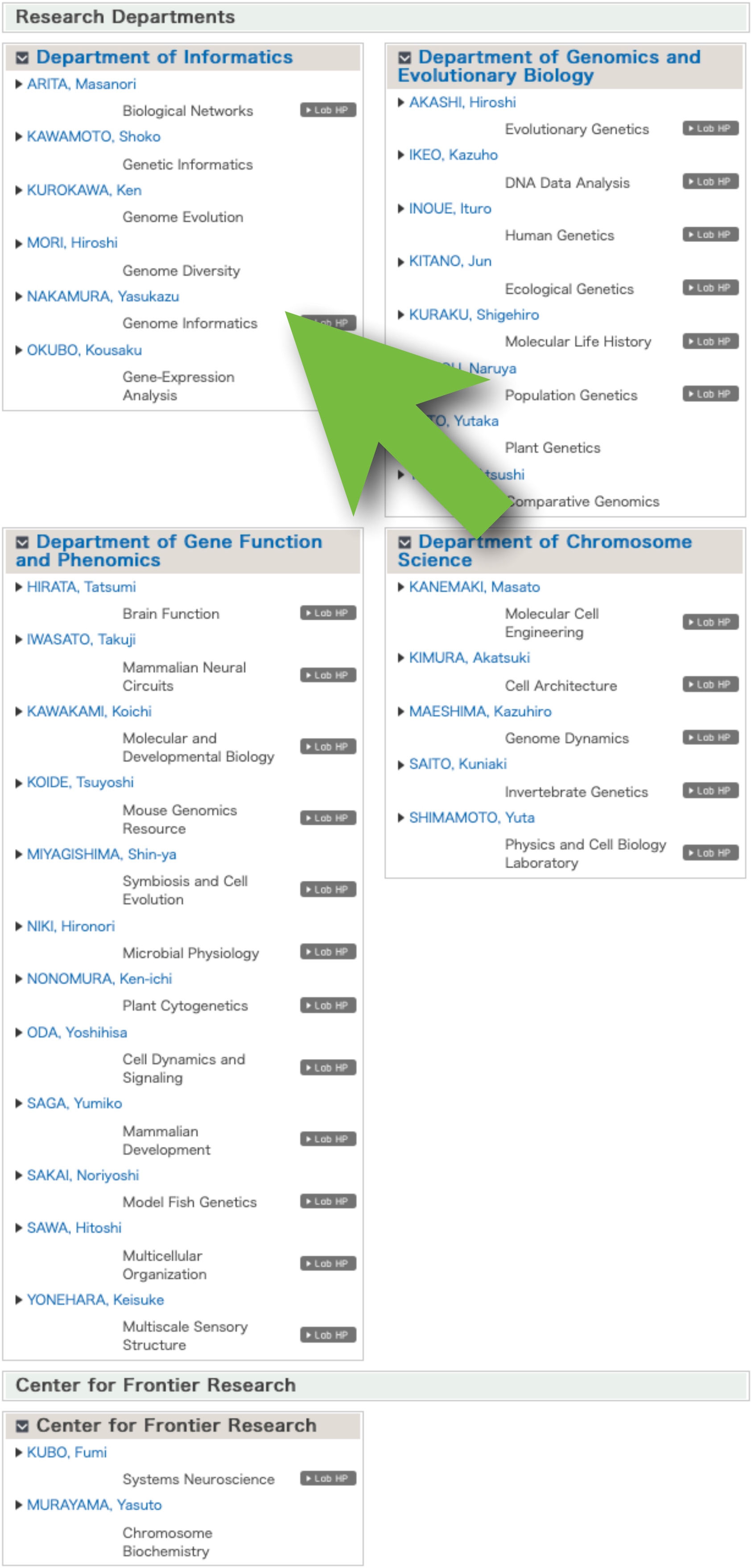
年前

https://www.youtube.com/user/ddbjvideo

Update from our lab

for Life and Health Sciences

National Institute of Genetics: www.nig.ac.jp



Research Departments

☑ Department of Informatics

- ARITA, Masanori
  - Biological Networks — Lab HP
- KAWAMOTO, Shoko
  - Genetic Informatics
- KUROKAWA, Ken
  - Genome Evolution
- MORI, Hiroshi
  - Genome Diversity
- NAKAMURA, Yasukazu
  - Genome Informatics — Lab HP
- OKUBO, Kousaku
  - Gene-Expression Analysis

☑ Department of Genomics and Evolutionary Biology

- AKASHI, Hiroshi
  - Evolutionary Genetics — Lab HP
- IKEO, Kazuho
  - DNA Data Analysis — Lab HP
- INOUE, Ituro
  - Human Genetics — Lab HP
- KITANO, Jun
  - Ecological Genetics — Lab HP
- KURAKU, Shigehiro
  - Molecular Life History — Lab HP
- U, Naruya
  - Population Genetics — Lab HP
- TO, Yutaka
  - Plant Genetics
- tsushi
  - omparative Genomics

☑ Department of Gene Function and Phenomics

- HIRATA, Tatsumi
  - Brain Function — Lab HP
- IWASATO, Takuji
  - Mammalian Neural Circuits — Lab HP
- KAWAKAMI, Koichi
  - Molecular and Developmental Biology — Lab HP
- KOIDE, Tsuyoshi
  - Mouse Genomics Resource — Lab HP
- MIYAGISHIMA, Shin-ya
  - Symbiosis and Cell Evolution — Lab HP
- NIKI, Hironori
  - Microbial Physiology — Lab HP
- NONOMURA, Ken-ichi
  - Plant Cytogenetics — Lab HP
- ODA, Yoshihisa
  - Cell Dynamics and Signaling — Lab HP
- SAGA, Yumiko
  - Mammalian Development — Lab HP
- SAKAI, Noriyoshi
  - Model Fish Genetics — Lab HP
- SAWA, Hitoshi
  - Multicellular Organization — Lab HP
- YONEHARA, Keisuke
  - Multiscale Sensory Structure — Lab HP

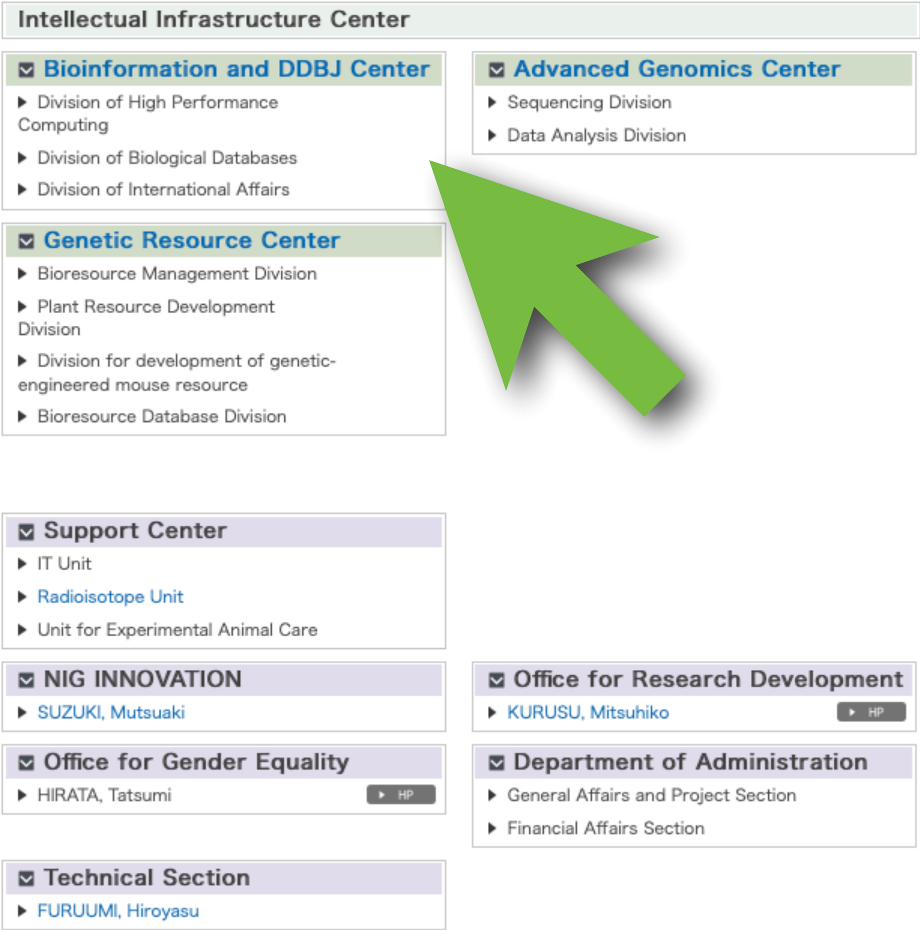
☑ Department of Chromosome Science

- KANEMAKI, Masato
  - Molecular Cell Engineering — Lab HP
- KIMURA, Akatsuki
  - Cell Architecture — Lab HP
- MAESHIMA, Kazuhiro
  - Genome Dynamics — Lab HP
- SAITO, Kuniaki
  - Invertebrate Genetics — Lab HP
- SHIMAMOTO, Yuta
  - Physics and Cell Biology Laboratory — Lab HP

Center for Frontier Research

☑ Center for Frontier Research

- KUBO, Fumi
  - Systems Neuroscience — Lab HP
- MURAYAMA, Yasuto
  - Chromosome Biochemistry



Intellectual Infrastructure Center

☑ Bioinformation and DDBJ Center

- Division of High Performance Computing
- Division of Biological Databases
- Division of International Affairs

☑ Advanced Genomics Center

- Sequencing Division
- Data Analysis Division

☑ Genetic Resource Center

- Bioresource Management Division
- Plant Resource Development Division
- Division for development of genetic-engineered mouse resource
- Bioresource Database Division

☑ Support Center

- IT Unit
- Radioisotope Unit
- Unit for Experimental Animal Care

☑ NIG INNOVATION

- SUZUKI, Mutsuaki

☑ Office for Research Development

- KURUSU, Mitsuhiko — HP

☑ Office for Gender Equality

- HIRATA, Tatsumi — HP

☑ Department of Administration

- General Affairs and Project Section
- Financial Affairs Section

☑ Technical Section

- FURUUMI, Hiroyasu

NATIONAL INSTITUTE OF GENETICS · JAPAN ·

SOKENDAI

Genome Informatics lab.

- Yaz Nakamura, Professor
- Yasuhiro Tanizawa, Assist. Professor
- Takako Mochizuki, Project Researcher
- Mika Sakamoto, Project Researcher
- Satomi Asano, Project Researcher
- Naoko Sakamoto, Technical Assistant
- Fumi Hayashi, Technical Assistant
- Misato Godo, Technical Assistant
- Dauyey Kaisar, D5 Student [coral fish evolution]
- Hanjie Mao, D5 Student [cat genetic diseases]
- Mohamed Elmanzalawi, D1 Student [human rare diseases]
- Takatomo Fujisawa, Researcher (DDBJ)

Nakamura lab’s genome works

Citrus unshiu

An orange

農研機構
NARO 国立研究開発法人 農業・食品産業技術総合研究機構

Gryllus bimaculatus

A cricket

東京大学
THE UNIVERSITY OF TOKYO

京都大学
KYOTO UNIVERSITY

MONASH University

GMI
GREGOR MENDEL INSTITUTE OF MOLECULAR PLANT BIOLOGY

KINDAI UNIVERSITY

NIBB

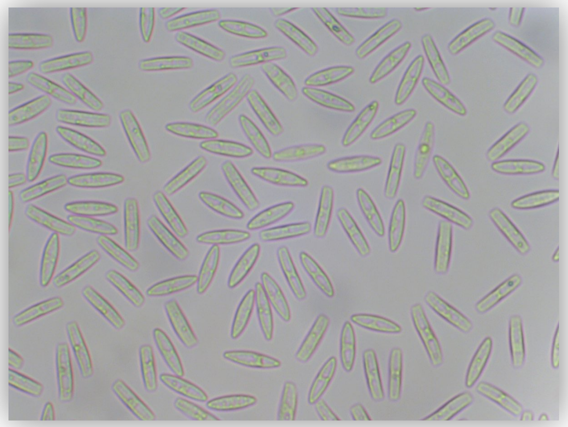
et al.

Marchantia polymorpha

A liverwort

Nitzschia spp.

A non-photosynthetic diatom



京都大学
KYOTO UNIVERSITY

筑波大学
University of Tsukuba

国立科学博物館
National Museum of Nature and Science

Felis catus

anicom medical
アニコム先進医療研究所株式会社

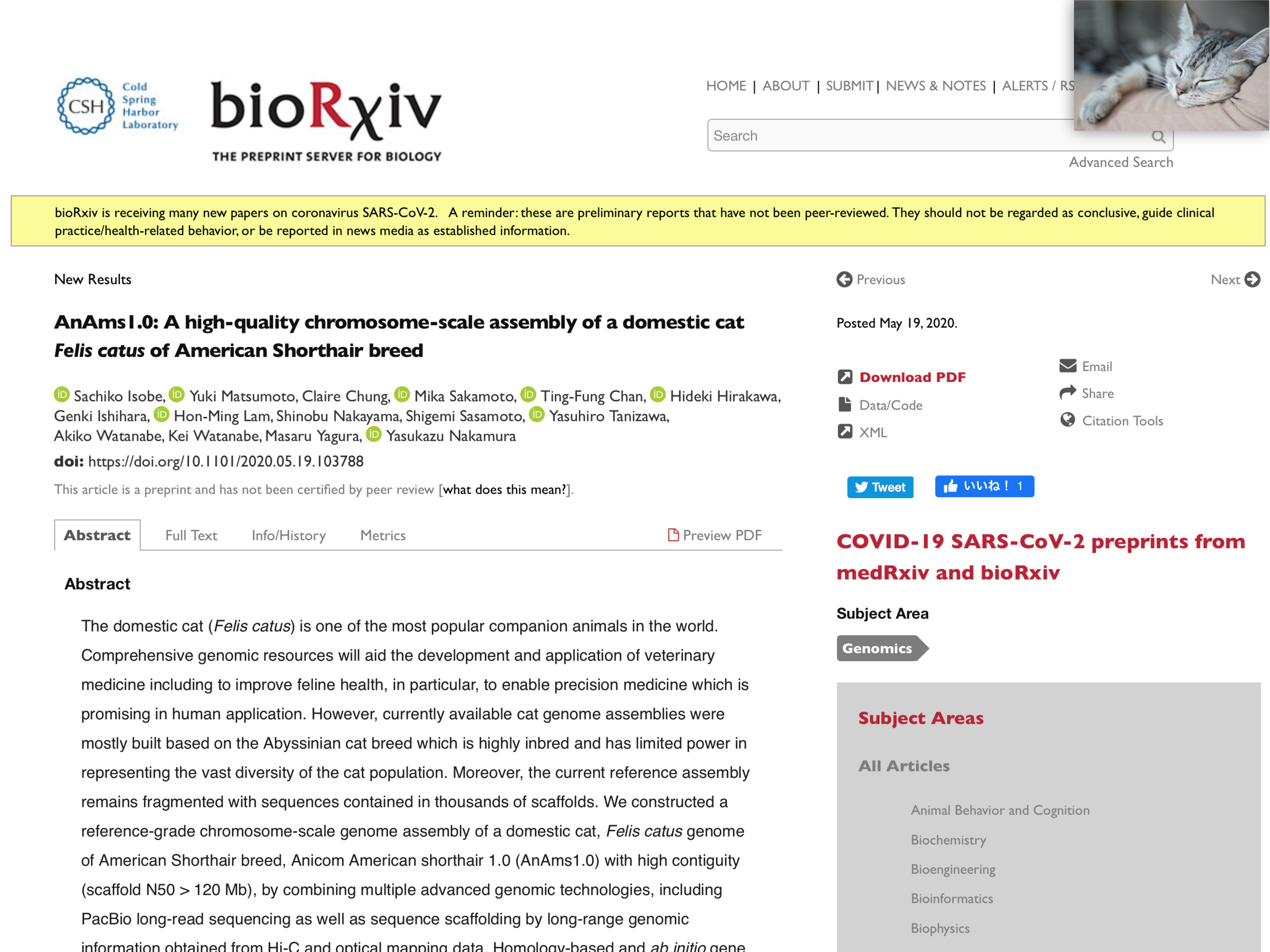
公益財団法人
かずさDNA研究所

香港中文大學
The Chinese University of Hong Kong

Tea tree (2,500)

Wasabi (100)

Strawberry (400)



New Results

Previous Next

AnAms1.0: A high-quality chromosome-scale assembly of a domestic cat *Felis catus* of American Shorthair breed

Sachiko Isobe, Yuki Matsumoto, Claire Chung, Mika Sakamoto, Ting-Fung Chan, Hideki Hirakawa, Genki Ishihara, Hon-Ming Lam, Shinobu Nakayama, Shigemi Sasamoto, Yasuhiro Tanizawa, Akiko Watanabe, Kei Watanabe, Masaru Yagura, Yasukazu Nakamura

doi: https://doi.org/10.1101/2020.05.19.103788

Posted May 19, 2020.

Abstract

The domestic cat (*Felis catus*) is one of the most popular companion animals in the world. Comprehensive genomic resources will aid the development and application of veterinary medicine including to improve feline health, in particular, to enable precision medicine which is promising in human application. However, currently available cat genome assemblies were mostly built based on the Abyssinian cat breed which is highly inbred and has limited power in representing the vast diversity of the cat population. Moreover, the current reference assembly remains fragmented with sequences contained in thousands of scaffolds. We constructed a reference-grade chromosome-scale genome assembly of a domestic cat, *Felis catus* genome of American Shorthair breed, Anicom American shorthair 1.0 (AnAms1.0) with high contiguity (scaffold N50 > 120 Mb), by combining multiple advanced genomic technologies, including PacBio long-read sequencing as well as sequence scaffolding by long-range genomic information obtained from Hi-C and optical mapping data. Homology-based and *ab initio* gene

COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv

Subject Area

Genomics

Subject Areas

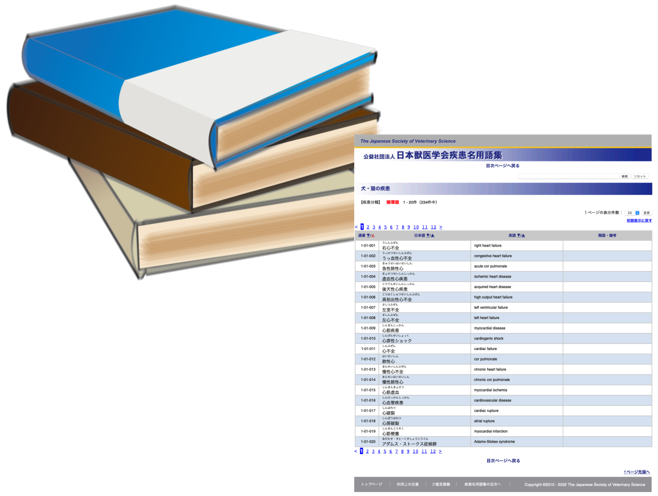
All Articles

Animal Behavior and Cognition
Biochemistry
Bioengineering
Bioinformatics
Biophysics

Table 3. Statistics of the AnAms1.0 and felcat genomes.

	AnAms1.0		felCat9	
	All scaffolds	Scaffolds correspond to nucleic chromosomes	All scaffolds	Scaffolds correspond to nucleic chromosomes
No. of sequences	20	19	4,508	19
Total length, bp	2,493,141,6[illegible]	2,493,140,262	2,521,863[illegible]5	2,460,251,910
Avg. length, bp	124,657[illegible]	131,217,909	55[illegible]	129,486,943
Max. length, bp	243,50[illegible]	243,504,654	242,[illegible]	242,100,913
Min. length, bp	1,3[illegible]	41,750,578	1[illegible]9	44,648,284
N50 length, bp	151,107,6[illegible]	151,107,676	149,751[illegible]9	149,751,809
A	718,322,616	718,322,255	721,302,099	703,154,254
T	719,032,844	719,032,444	721,735,169	703,748,672
G	517,632,441	517,632,143	516,811,973	504,031,439
C	517,568,802	517,568,480	516,603,929	503,993,730
N	0.83	0.83	1.8	1.84
Total, ATGC	2,472,556,703	2,472,555,322	2,476,453,170	2,414,928,095
GC%, ATGC	41.9	41.9	41.7	41.7

Expansion of Cats-I for vet information infrastructure



The Japanese Society of Veterinary Science

公益社団法人 日本獣医学会疾患名用語集

目次ページへ戻る

検索 リセット

犬・猫の疾患

【疾患分類】 循環器 1 - 20件（234件中）

1ページの表示件数： 20 変更

初期表示に戻す

< 1 2 3 4 5 6 7 8 9 10 11 12 >

通番 ▼/▲	日本語 ▼/▲	英語 ▼/▲	類語・備考
1-01-001	うしんふぜん 右心不全	right heart failure	
1-01-002	うっけつせいしんふぜん うっ血性心不全	congestive heart failure	
1-01-003	きゅうせいはいせいしん 急性肺性心	acute cor pulmonale	
1-01-004	きょけつせいしんしっかん 虚血性心疾患	ischemic heart disease	
1-01-005	こうてんせいしんしっかん 後天性心疾患	acquired heart disease	
1-01-006	こうはくしゅつせいしんふぜん 高拍出性心不全	high output heart failure	
1-01-007	さしつふぜん 左室不全	left ventricular failure	
1-01-008	さしんふぜん 左心不全	left heart failure	
1-01-009	しんきんしっかん 心筋疾患	myocardial disease	
1-01-010	しんげんせいしょっく 心原性ショック	cardiogenic shock	
1-01-011	しんふぜん 心不全	cardiac failure	
1-01-012	はいせいしん 肺性心	cor pulmonale	
1-01-013	まんせいしんふぜん 慢性心不全	chronic heart failure	
1-01-014	まんせいはいせいしん 慢性肺性心	chronic cor pulmonale	
1-01-015	しんきんきょけつ 心筋虚血	myocardial ischemia	
1-01-016	しんけっかんしっかん 心血管疾患	cardiovascular disease	
1-01-017	しんはれつ 心破裂	cardiac rupture	
1-01-018	しんぼうはれつ 心房破裂	atrial rupture	
1-01-019	しんきんこうそく 心筋梗塞	myocardial infarction	
1-01-020	あだむす・すとーくすしょうこうぐん アダムス・ストークス症候群	Adams-Stokes syndrome	

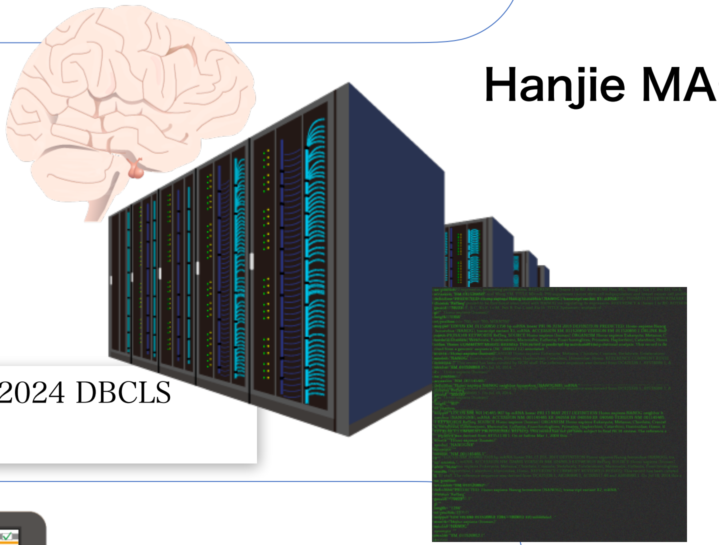
< 1 2 3 4 5 6 7 8 9 10 11 12 >

目次ページへ戻る

↑ページ先頭へ

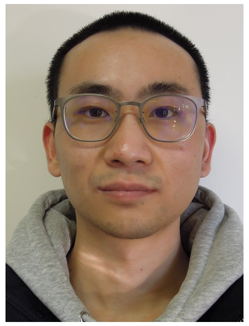
トップページ | 利用上の注意 | ご意見登録 | 疾患名用語集の目次へ |

1. publication curation

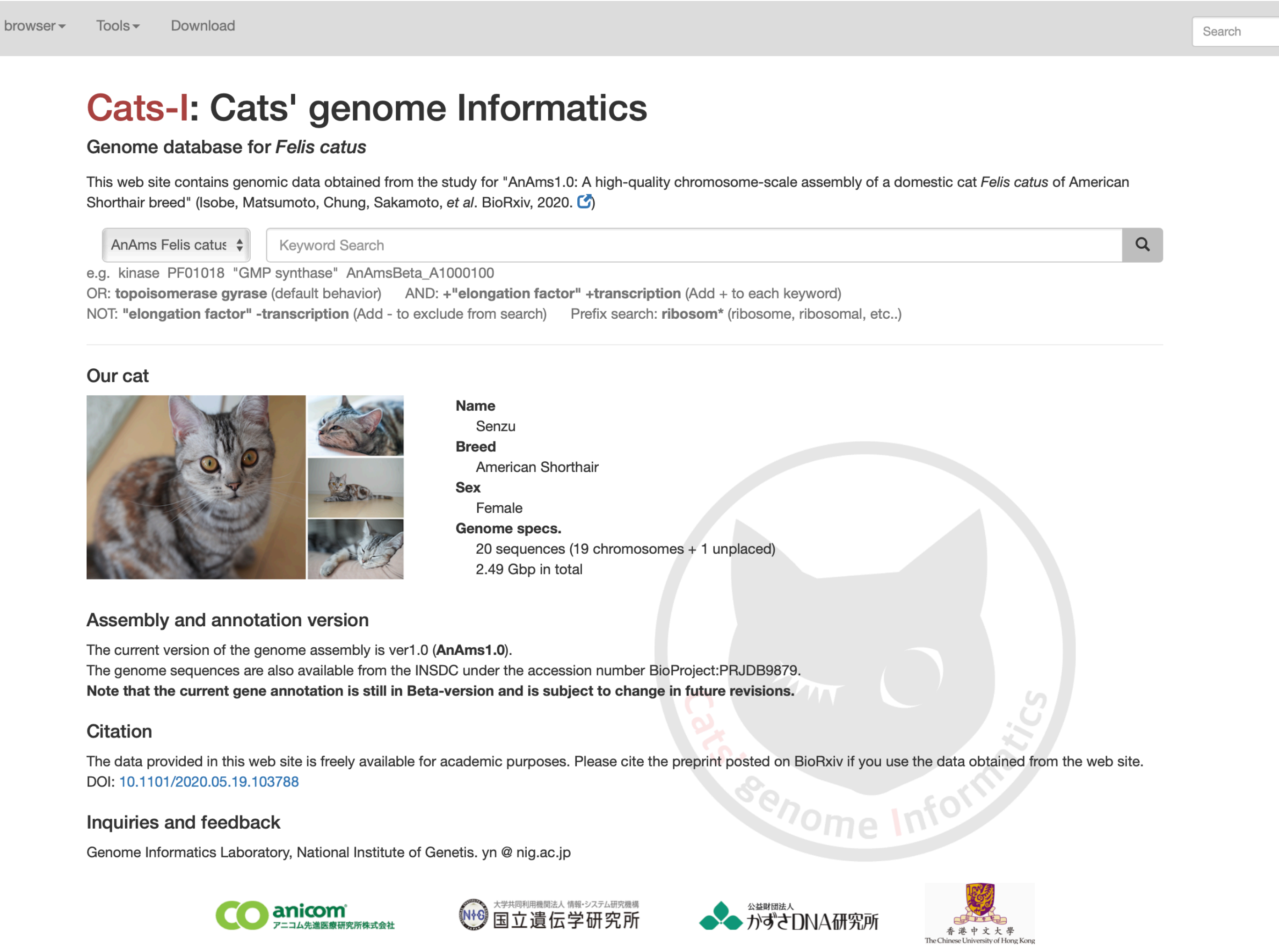


Togo picture gallery © 2024 DBCLS
TogoTV / CC-BY-4.0

Hanjie MAO



2. variation finding from SRA



Cats-I: Cats' genome Informatics

Genome database for *Felis catus*

This web site contains genomic data obtained from the study for "AnAms1.0: A high-quality chromosome-scale assembly of a domestic cat *Felis catus* of American Shorthair breed" (Isobe, Matsumoto, Chung, Sakamoto, *et al.* BioRxiv, 2020.)

AnAms Felis catus | Keyword Search

e.g. kinase PF01018 "GMP synthase" AnAmsBeta_A1000100
OR: **topoisomerase gyrase** (default behavior) AND: **+"elongation factor" +transcription** (Add + to each keyword)
NOT: **"elongation factor" -transcription** (Add - to exclude from search) Prefix search: **ribosom*** (ribosome, ribosomal, etc..)

Our cat

Name
Senzu
Breed
American Shorthair
Sex
Female
Genome specs.
20 sequences (19 chromosomes + 1 unplaced)
2.49 Gbp in total

Assembly and annotation version

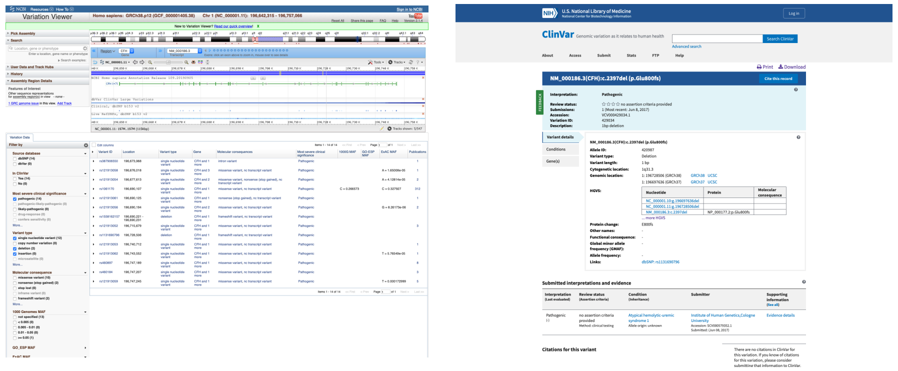
The current version of the genome assembly is ver1.0 (**AnAms1.0**).
The genome sequences are also available from the INSDC under the accession number BioProject:PRJDB9879.
Note that the current gene annotation is still in Beta-version and is subject to change in future revisions.

Citation

The data provided in this web site is freely available for academic purposes. Please cite the preprint posted on BioRxiv if you use the data obtained from the web site.
DOI: 10.1101/2020.05.19.103788

Inquiries and feedback

Genome Informatics Laboratory, National Institute of Genetis. yn @ nig.ac.jp

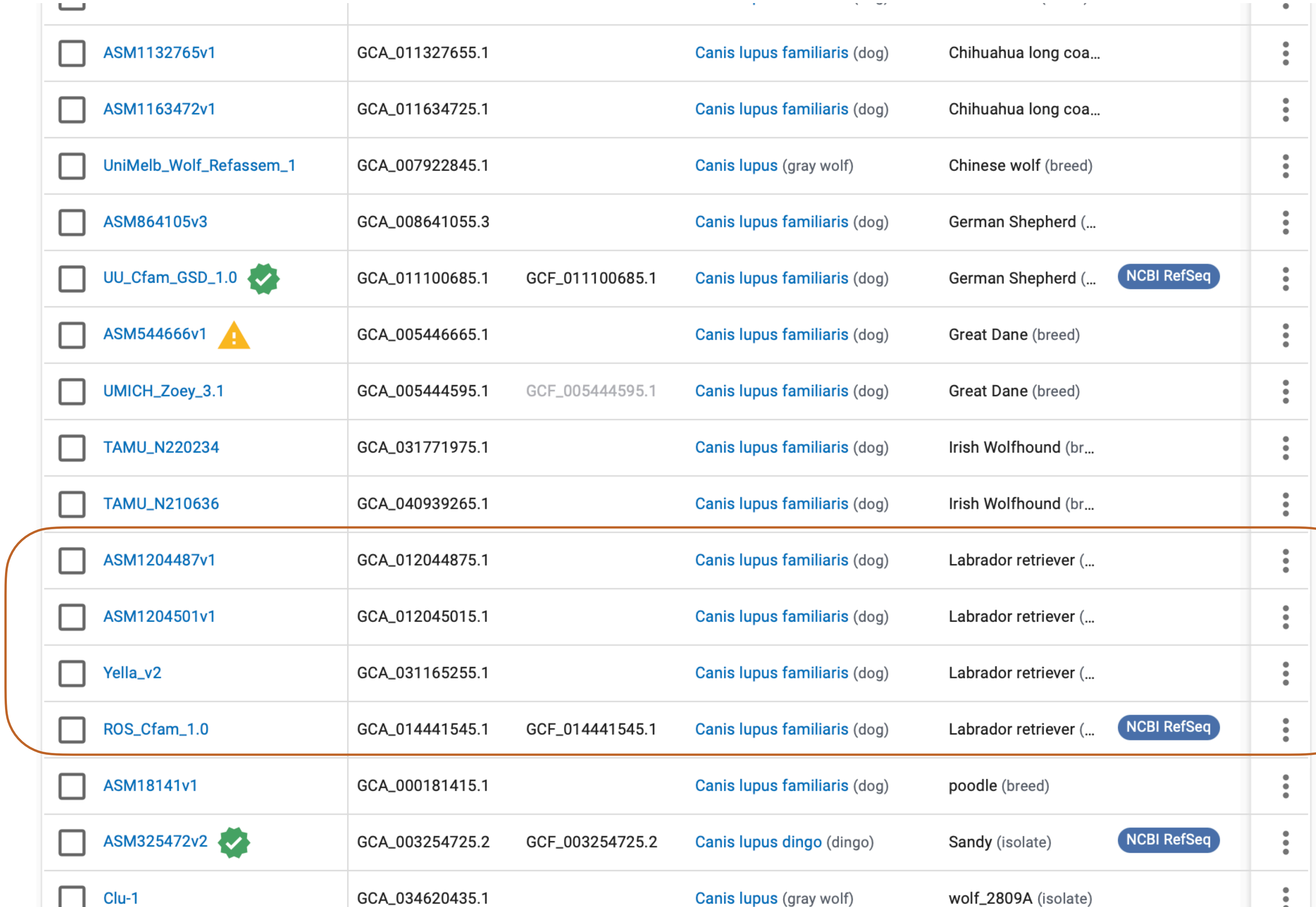


Cat version of “ClinVar” and more

Complete genome project for Big dogs [new]

- 38 chr + XY, 2.4 Gbps
- For Dogs and Human Healthcare
- <u>Aiming to make T2T</u> (376 scaffolds ⇒ 39!)
    - Pacbio Revio Hifi reads

Canis lupus



☐	ASM1132765v1	GCA_011327655.1		Canis lupus familiaris (dog)	Chihuahua long coa…		⋮
☐	ASM1163472v1	GCA_011634725.1		Canis lupus familiaris (dog)	Chihuahua long coa…		⋮
☐	UniMelb_Wolf_Refassem_1	GCA_007922845.1		Canis lupus (gray wolf)	Chinese wolf (breed)		⋮
☐	ASM864105v3	GCA_008641055.3		Canis lupus familiaris (dog)	German Shepherd (…		⋮
☐	UU_Cfam_GSD_1.0 ✓	GCA_011100685.1	GCF_011100685.1	Canis lupus familiaris (dog)	German Shepherd (…	NCBI RefSeq	⋮
☐	ASM544666v1 ⚠	GCA_005446665.1		Canis lupus familiaris (dog)	Great Dane (breed)		⋮
☐	UMICH_Zoey_3.1	GCA_005444595.1	GCF_005444595.1	Canis lupus familiaris (dog)	Great Dane (breed)		⋮
☐	TAMU_N220234	GCA_031771975.1		Canis lupus familiaris (dog)	Irish Wolfhound (br…		⋮
☐	TAMU_N210636	GCA_040939265.1		Canis lupus familiaris (dog)	Irish Wolfhound (br…		⋮
☐	ASM1204487v1	GCA_012044875.1		Canis lupus familiaris (dog)	Labrador retriever (…		⋮
☐	ASM1204501v1	GCA_012045015.1		Canis lupus familiaris (dog)	Labrador retriever (…		⋮
☐	Yella_v2	GCA_031165255.1		Canis lupus familiaris (dog)	Labrador retriever (…		⋮
☐	ROS_Cfam_1.0	GCA_014441545.1	GCF_014441545.1	Canis lupus familiaris (dog)	Labrador retriever (…	NCBI RefSeq	⋮
☐	ASM18141v1	GCA_000181415.1		Canis lupus familiaris (dog)	poodle (breed)		⋮
☐	ASM325472v2 ✓	GCA_003254725.2	GCF_003254725.2	Canis lupus dingo (dingo)	Sandy (isolate)	NCBI RefSeq	⋮
☐	Clu-1	GCA_034620435.1		Canis lupus (gray wolf)	wolf_2809A (isolate)		⋮

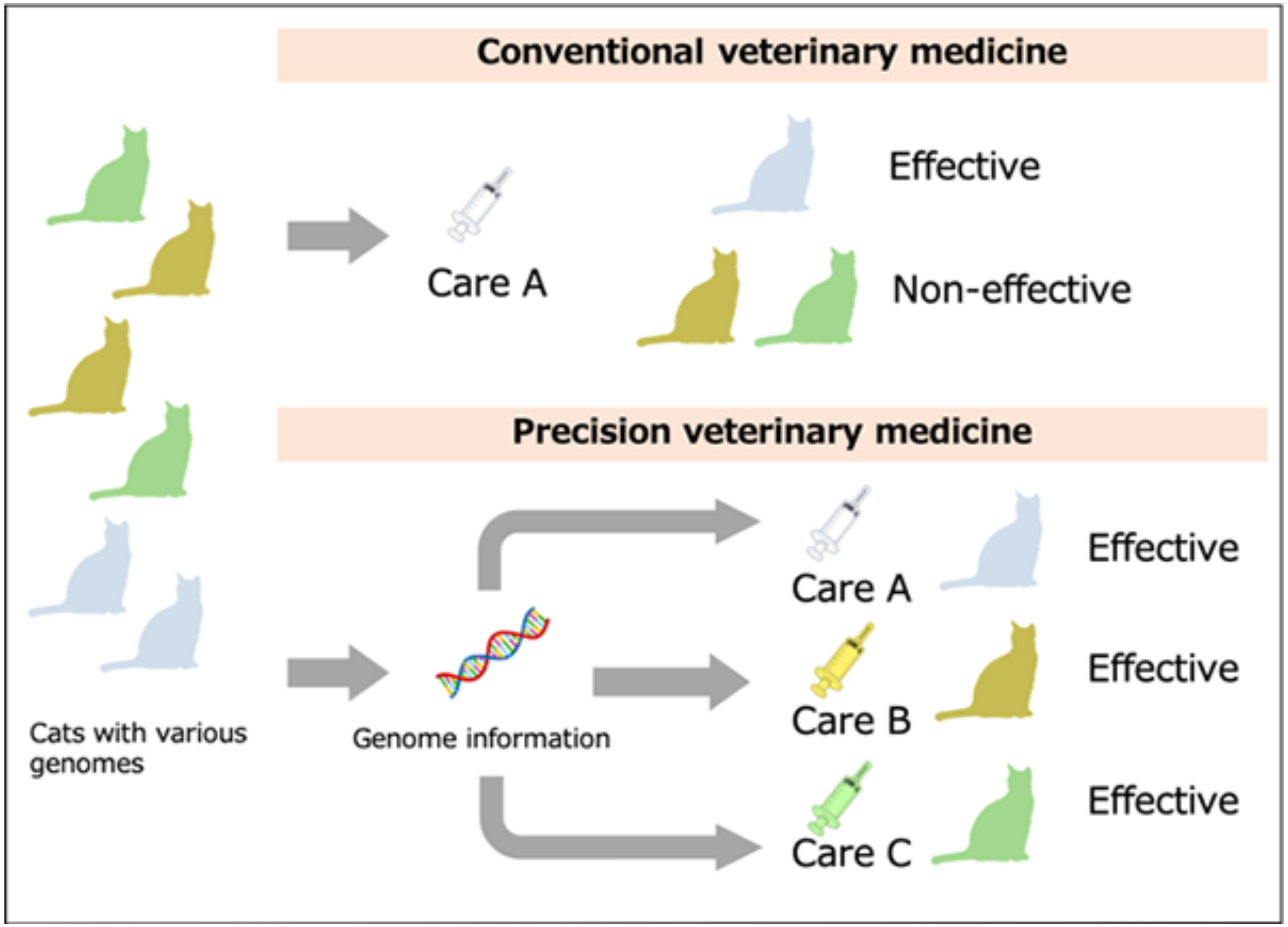
COMIT One Medicine 創薬シーズ開発・育成研究教育拠点
Center for One Medicine Innovative Translational Research

岐阜大学
GIFU UNIVERSITY

名古屋大学
NAGOYA UNIVERSITY

大学共同利用機関法人 情報・システム研究機構
国立遺伝学研究所

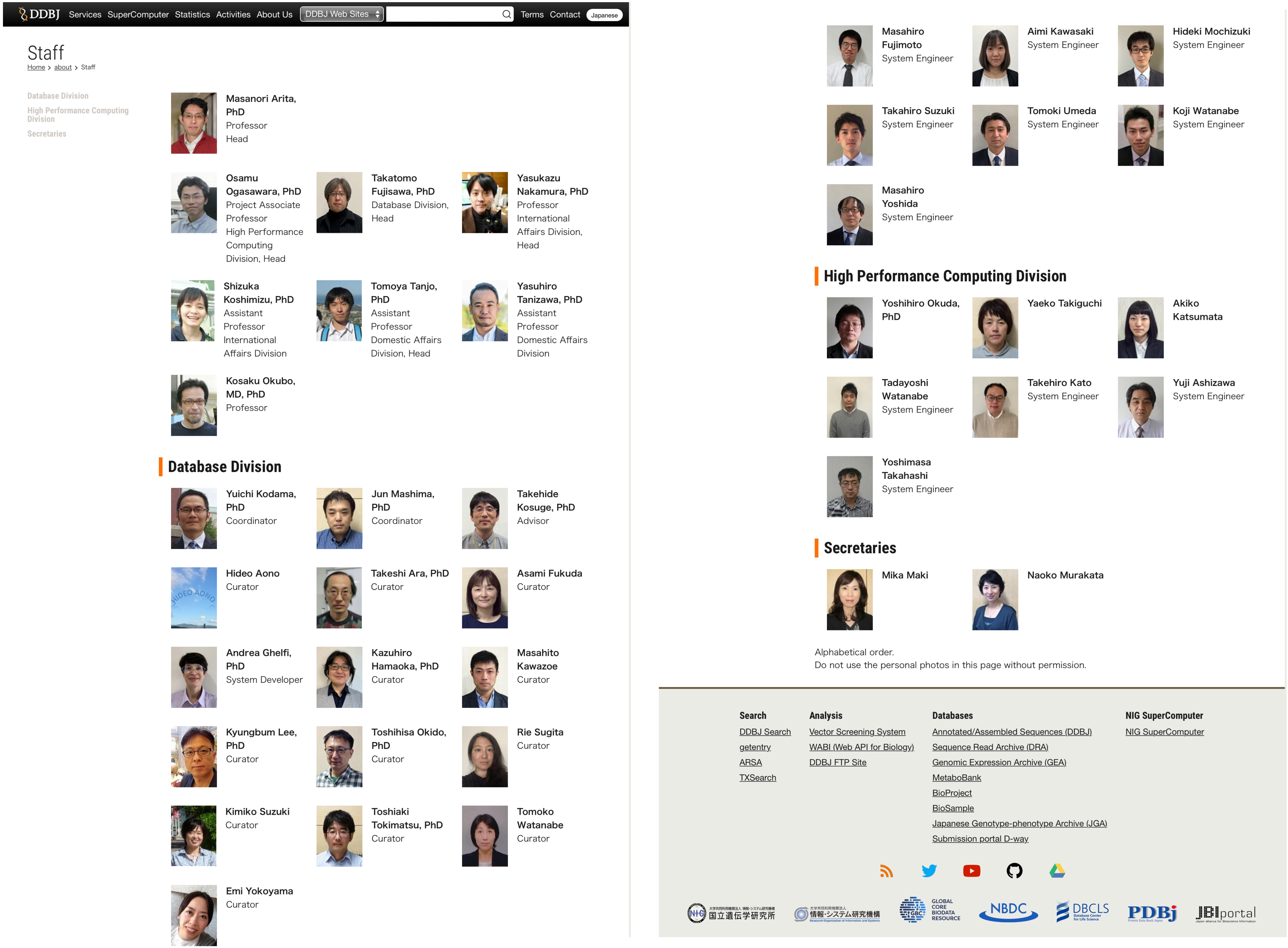
precision medicine for companion animals



“Personal” Medicine for companion animals

In traditional veterinary medicine, the same treatment is given to any individual where some cats are not responsive to specific treatments. With the genomic approach, individually suited treatments can be applied based on individual traits predicted from genome information for more effective and efficient veterinary medicine.

Acknowledgments: DDBJ members



Staff

Home > about > Staff

Database Division
High Performance Computing Division
Secretaries

Masanori Arita, PhD
Professor
Head

Osamu Ogasawara, PhD
Project Associate Professor
High Performance Computing Division, Head

Takatomo Fujisawa, PhD
Database Division, Head

Yasukazu Nakamura, PhD
Professor
International Affairs Division, Head

Shizuka Koshimizu, PhD
Assistant Professor
International Affairs Division

Tomoya Tanjo, PhD
Assistant Professor
Domestic Affairs Division, Head

Yasuhiro Tanizawa, PhD
Assistant Professor
Domestic Affairs Division

Kosaku Okubo, MD, PhD
Professor

Database Division

Yuichi Kodama, PhD
Coordinator

Jun Mashima, PhD
Coordinator

Takehide Kosuge, PhD
Advisor

Hideo Aono
Curator

Takeshi Ara, PhD
Curator

Asami Fukuda
Curator

Andrea Ghelfi, PhD
System Developer

Kazuhiro Hamaoka, PhD
Curator

Masahito Kawazoe
Curator

Kyungbum Lee, PhD
Curator

Toshihisa Okido, PhD
Curator

Rie Sugita
Curator

Kimiko Suzuki
Curator

Toshiaki Tokimatsu, PhD
Curator

Tomoko Watanabe
Curator

Emi Yokoyama
Curator

Masahiro Fujimoto
System Engineer

Aimi Kawasaki
System Engineer

Hideki Mochizuki
System Engineer

Takahiro Suzuki
System Engineer

Tomoki Umeda
System Engineer

Koji Watanabe
System Engineer

Masahiro Yoshida
System Engineer

High Performance Computing Division

Yoshihiro Okuda, PhD

Yaeko Takiguchi

Akiko Katsumata

Tadayoshi Watanabe
System Engineer

Takehiro Kato
System Engineer

Yuji Ashizawa
System Engineer

Yoshimasa Takahashi
System Engineer

Secretaries

Mika Maki

Naoko Murakata

Alphabetical order.
Do not use the personal photos in this page without permission.

Search
- DDBJ Search
- getentry
- ARSA
- TXSearch

Analysis
- Vector Screening System
- WABI (Web API for Biology)
- DDBJ FTP Site

Databases
- Annotated/Assembled Sequences (DDBJ)
- Sequence Read Archive (DRA)
- Genomic Expression Archive (GEA)
- MetaboBank
- BioProject
- BioSample
- Japanese Genotype-phenotype Archive (JGA)
- Submission portal D-way

NIG SuperComputer
- NIG SuperComputer

https://www.ddbj.nig.ac.jp/about/staff-e.html

DDBJ
DNA Data Bank of Japan