International Vertebrate Genomes Project releases first 15 new genomes

Max Planck Society supports projects for high quality reference genomes

The international Vertebrate Genomes Project (VGP) is officially launched and releases 15 new reference genomes representing all five vertebrate classes – mammals, birds, reptiles, amphibians, and fish. These 15 genomes are the most complete versions of their species to date. The mission of the VGP is to sequence and assemble high quality, nearly error-free, and complete genomes of all 66,000 vertebrate species on Earth. The VGP data is currently being produced primarily by teams at three sequencing hubs: the Rockefeller University, USA, the Wellcome Sanger Institute, UK, and at the Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG) in Dresden, Germany. Two of the 15 released genomes, a bat and a fish, have been sequenced and assembled in Dresden.

With its ambitious mission the VGP aims to address fundamental questions in biology, conservation, and disease including identifying species most genetically at risk for extinction and preserving their genetic information for future generations. The high-quality VGP genomes will become the main references for their species and will be stored in the Genome Ark, a digital open-access library of genomes.

The current Phase 1 of the VGP – the VGP orders project - aims to create reference assemblies of selected species representing all 260 vertebrate orders that have diverged from each other shortly after the last mass extinction 66 million years ago. Studying these ordinal level species will help scientists determine what type of species survived the previous extinction event that wiped out the dinosaurs. Those studies can also give insights into how other species could survive the current 6th mass extinction event and help identify genetic variants that might protect these species from total extinction. Amongst the 15 new genomes are critically endangered species like the platypus, and the Kakapo parrot. Other species include the zebra finch songbird and the Anna’s hummingbird, which like parrots, belong to the only three vocal learning bird orders among over 40 orders of birds. Also, two vocal learning bat species are part of this first data release.

To conduct the VGP, the umbrella G10K organization, from which the project arose, has convened over 150 experts from academia, industry, and government, from 12 countries, to develop high-resolution sequencing methods that both reduce costs and eliminate the errors that plague current reference genomes. Many current reference genomes are riddled with errors—parts of genes are missing, some are incorrectly assembled, and other genes are completely missing. Consequently, researchers are potentially working with incorrect gene sequences and structures hampering their genomic studies. The new VGP genomes eliminate most of these errors.

The MPI-CBG and in particular its bioinformatics researchers at the Center for Systems Biology Dresden (CSBD) is involved in the sequencing, assembly and annotation of the initial Phase I genomes of the VGP project with a focus on bats and fish. The Dresden scientists are part of the DRESDEN-concept Genome Center (DCGC) and have special expertise in using various long-read sequencing and long-range scaffolding technologies. The Dresden hub, led by Prof. Dr. Eugene Myers has contributed two genomes of the 15 released species: the greater horseshoe bat (Rhinolophus ferrumequinum) and the flier cichlid fish (Archocentrus centrarchus). In the future, about 10-20% of the VGP species are
expected to be sequenced in Dresden. Prof. Eugene Myers, director at the MPI-CBG and founder of the CSBD says, “The advances in long-read sequencing is revolutionizing DNA sequencing. After a 10-year hiatus, this trend inspired me to return to genome assembly as I believe it implies that we will ultimately be able to produce near-perfect genome reconstructions. I think this capability is going to dramatically alter the landscape of genomics.”

In addition to the VGP, the MPI-CBG and the CSBD are actively engaged in synergistic international sequencing projects. The Bat1K project has the goal of sequencing all 1,300 bat species, many of which live unusually long or have near-perfect immune systems. Six bat genomes will be released in the near future, and another 25 species are being prepared to study aging, immunity, and vocal-learning in collaboration with the BatIK consortium, which includes partners Sonja Vernes from the Max Planck Institute for Psycholinguistics in the Netherlands and Emma Teeling of the University College Dublin, UK. Another project is the Euro-Fish project, which aims to sequence almost all 600 species of fish swimming in European fresh waters. One of our main collaborators is Prof. Dr. Axel Meyer of the University of Konstanz. The Max Planck Society is funding the initial genomes from these synergistic projects. All the genomes will be sequenced to the high quality standard set by the VGP and will be placed in the Genome Ark repository, where one day all 66,000 vertebrates will be recorded.

The 15 genomes created through the VGP:
1. Mammals (4 species)
   - Two bat species, Greater horseshoe bat (Rhinolophus ferrumequinum) and Pale spear-nose bat (Phyllostomus discolor), used as models for longevity and vocal learning
   - The Canada lynx (Lynx canadensis), once nearly extinct in the United States and now recovering
   - The duck-billed platypus (Ornithorhynchus anatinus), an egg-laying mammal with reptilian traits

2. Reptiles (1 species)
   - A newly discovered turtle species from Mexico, Goode’s Thornscrub Tortoise (Gopherus evgoodei)

3. Amphibians (1 species)
   - Two-lined caecilian (Rhinatrema bivittatum), a limbless amphibian that resembles a snake

4. Birds (3 species, 4 genomes)
   - In addition to the kakapo (Strigops habroptilus), the VGP re-sequenced species from two other bird orders to represent the only three vocal learning birds among more than 40 avian orders
   - A male and female zebra finch (Taeniopygia guttata), the most commonly studied vocal learner
   - Anna’s hummingbird (Calypte anna), belonging to the smallest group of birds

5. Fish (5 species)
   - These species represent a large diversity of traits and are used to study species evolution and adaptation:
     - Flier Cichlid (Archenocentrus centrarchus), native to Central America
     - Eastern happy (Astatotilapia calliptera), also a cichlid fish Native to Lake Malawi, Africa
     - Climbing perch (Anabas testudineus), native to inland waters of Southeast Asia
     - Tire track eel (Mastacembelus armatus), native to rivers of Southeast Asia
     - Blunt-snouted clingfish (Gouania willdenowi), native to north Mediterranean coast, Syria to Spain

About the MPI-CBG
The Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG) is one of 84 institutes of the Max Planck Society, an independent, non-profit organization in Germany. 500 curiosity-driven scientists from over 50 countries ask: How do cells form tissues? The basic research programs of the MPI-CBG span multiple scales of magnitude, from molecular assemblies to organelles, cells, tissues, organs, and organisms.
About the CSBD
The Center for Systems Biology Dresden (CSBD) is a cooperation between the Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG), the Max Planck Institute for the Physics of Complex Systems (MPI-PKS) and the TU Dresden. The interdisciplinary center brings physicists, computer scientists, mathematicians and biologists together. The scientists develop theoretical and computational approaches to biological systems across different scales, from molecules to cells and from cells to tissues.

About the DRESDEN-concept Genome Center (DCGC)
The DCGC is a joint sequencing center between the Technische Universität Dresden and the MPI-CBG. It is one of four DFG-funded German competence centers for next generation sequencing. The cooperative project is an amalgamation of employees of the TU Dresden and MPI-CBG, as well as of the CSBD, and the Center for Regenerative Therapies Dresden (CRTD). The center consists of three platforms focusing on long read sequencing technologies, single cell sequencing, and short read sequencing.

About the Rockefeller University
The Rockefeller University is the world's leading biomedical research university and is dedicated to conducting innovative, high-quality research to improve the understanding of life for the benefit of humanity. Our 82 laboratories conduct research in neuroscience, immunology, biochemistry, genomics, and many other areas, and a community of 1,800 faculty, students, postdocs, technicians, clinicians, and administrative personnel work on our 14-acre Manhattan campus. Our unique approach to science has led to some of the world's most revolutionary and transformative contributions to biology and medicine. During Rockefeller's 117-year history, 25 of our scientists have won Nobel Prizes, 23 have won Albert Lasker Medical Research Awards, and 20 have garnered the National Medal of Science, the highest science award given by the United States.

About the Wellcome Sanger Institute
The Wellcome Sanger Institute is one of the world’s leading genome centres. Through its ability to conduct research at scale, it is able to engage in bold and long-term exploratory projects that are designed to influence and empower medical science globally. Institute research findings, generated through its own research programmes and through its leading role in international consortia, are being used to develop new diagnostics and treatments for human disease. To celebrate its 25th year in 2018, the Institute is sequencing 25 new genomes of species in the UK. Find out more at www.sanger.ac.uk or follow on Twitter @sangerinstitute.

The Vertebrate Genome Laboratory (VGL) at the Rockefeller University is a Resource Center specializing in ultra-High-Molecular Weight DNA (uHMW DNA) and long-read genomic technologies. The primary objective of the VGL is to generate at least one high-quality, phased, chromosome-level, annotated, reference genome assembly of all approximately 66,000 vertebrate species for the Vertebrate Genomes Project (G10K-VGP). The team is composed of four members, including a Director, two Research Support Specialists/Associates, and a Research Assistant. The VGL is equipped with four state of the art Pacific Biosciences Sequel™ sequencers, one Bionano Genomics Saphyr™ optical mapper, one 10× Genomics Chromium™ microfluidics platform, and all the necessary ancillary instruments for preparing uHMW DNA.

wissenschaftliche Ansprechpartner:
Prof. Eugene Myers
+49 (0) 351 210 1900
myers@mpi-cbg.de
Spix’s disk-winged bat (Thyroptera tricolor)
Sébastien Puechmaille