India's Genomic Data Set

For Prelims: GenomeIndia Project, Whole genome sequencing, Department of Biotechnology, Biotech-PRIDE Guidelines, IndiGen Project

For Mains: BioE3, India's Biotechnology Development and Innovations, Bioeconomy, Genome Sequencing

Source: PIB

Why in News?

India achieved a historic milestone at the **Genome India Data Conclave** in New Delhi with the launch of the **Indian Genomic Data (IGD) Set** under the **GenomeIndia Project (GIP) and** frameworks like the Framework for Exchange of Data Protocols **(FeED)** and the **Indian Biological Data Centre (IBDC) Portals**.

- These initiatives position India as a leader in **genomics,** providing global researchers access to **genome samples** and achieving self-reliance in genomic data.

What are the Key Highlights of the Genome India Data Conclave?

- **Indian Genomic Data Set:** A comprehensive Indian genomic data set was launched, featuring **10,000 whole genome sequencing (WGS)** samples, archived at the **Indian Biological Data Centre (IBDC), India's first national repository for life science data.**
  - The dataset is now accessible to researchers worldwide, supporting advancements in genomics research and personalized medicine.
  - The IBDC Portal facilitates seamless access to genetic data.
- **FeED Protocols:** The FeED ensures **ethical, transparent, and secure sharing of high-quality genomic data** under the **Biotech-PRIDE (Promotion of Research and Innovation through Data Exchange) Guidelines.**
- **GenomeIndia Project**: Emphasized the significance of the **Genome India Project (GIP),** led by the **Department of Biotechnology (DBT),** this initiative creates a comprehensive database of India's genetic diversity.

Biotech-PRIDE Guidelines

- The "Biotech-PRIDE Guidelines," released in 2021 by DBT, enable the exchange of **biological data across research groups in India**.
  - They provide a framework for sharing knowledge, ensuring better integration, decision-making, and equitable access.
  - They promote timely sharing, avoid duplication, and maximize the benefits of public investment in research.
- The guidelines will be implemented through the **Indian Biological Data Centre (IBDC)** at the **Regional Centre for Biotechnology (RCB), Haryana.**

- Under the Biotech-PRIDE Guidelines, existing datasets will be connected to the IBDC, forming the **Bio-Grid.**
    - This Bio-Grid will serve as a **National Repository for biological data,** enabling its exchange while ensuring safety, standards, and quality, and establishing clear data access protocols.
- The RCB, established by the DBT, focuses on biotechnology education, research, and training.
    - It is recognized as an **Institution of National Importance in 2016,** RCB fosters innovation and develops skilled human resources to address challenges in health, agriculture, and the environment.

What is the GenomeIndia Project?

- **About:** The GIP is a flagship initiative launched in 2020 by the DBT, with the objective of **mapping the genetic diversity of India.**
    - It aims to **sequence and analyze the genomes** of India's diverse population groups, providing insights into the country's unique genetic makeup.
- **Objective:** Establish a **baseline genetic map** to study health, disease predisposition, and population-specific traits.
- **Scope:** The first phase of GIP involves sequencing the **genomes of 10,000 individuals across 99 ethnic groups.** Long-term plans aim to scale this to **1 million genomes.**
    - The second phase of the GIP focuses on sequencing genomes of individuals with **specific diseases like cancer, diabetes, and rare diseases.**
        - This will help identify genes linked to these conditions by comparing diseased genomes with healthy ones.
- **Significance for India:** With over **4,600 distinct population groups,** India's genetic diversity is unparalleled.
    - The project aims to uncover **genetic factors unique to Indian populations,** such as rare diseases and **mutations like MYBPC3,** linked to early cardiac arrest, which are not found in **global databases.**

Genome Sequencing

- **Deoxyribonucleic Acid:** The DNA is a molecule that carries **genetic information**. It is the blueprint for the growth, development, and functioning of all living organisms.
    - DNA is a two-stranded molecule with a twisted shape, known as a **double helix.**
    - Each strand of DNA is composed of **nucleotides,** which include a **phosphate molecule,** a deoxyribose sugar, and a **nitrogen-containing base.**
- **Genome:** The genome is the complete set of **DNA instructions in a cell.** In humans, it consists of **23 pairs of chromosomes**.
    - One copy of the human genome contains approximately **3 billion base pairs of DNA,** distributed across these **23 chromosomes.**
    - The genome holds all the information necessary for an individual's development and functioning.
- **Genes:** The **basic units of heredity** passed from parent to child. They are made up of **DNA sequences** and are arranged at specific locations on chromosomes within the cell's nucleus.
- **Genome Sequencing:** It involves deciphering the order of the **nucleotide bases** (**adenine (A), cytosine (C), guanine (G), and thymine (T))** in the **Deoxyribonucleic Acid (DNA).**
    - This process helps in understanding the genetic makeup of an individual, revealing insights into **traits, health risks, and potential diseases.**
    - Genome Sequencing can be focused on a particular gene, region, or small section of the genome.
- **Whole Genome Sequencing (WGS):** Involves sequencing the entire genome of an organism in one go, including all of **its genes and non-coding regions** (the entire DNA sequence).
    - **WGS** provides a complete and comprehensive map of an organism's genetic material.

// › Difference Between Gene Editing and Gene Sequencing:

Characteristics	Gene Sequencing	Gene Editing
Definition	The process of determining the precise order of nucleotides (A, T, C, G) in a DNA or RNA molecule.	The process of making targeted modifications to the DNA sequence of a gene or genes.
Purpose	To obtain the complete or partial sequence of a gene, a set of genes, or an entire genome.	To introduce desired changes, such as correcting genetic defects, modifying gene expression, or introducing new genetic traits.
Techniques	Sanger sequencing, Next-Generation Sequencing (NGS), and others.	CRISPR-Cas9, zinc finger nucleases, TALENs, and other specialised tools.
Outcome	Provides information about the genetic makeup and composition of an organism.	Allows for the direct manipulation and alteration of the genetic code.
Modification	Does not directly modify the genetic material.	Enables the addition, removal, or alteration of specific DNA sequences.

What are the Key Benefits of Indigenous Genomic Data?

- **Personalized Medicine**: The healthcare system in India faces challenges due to a **one-size-fits-all approach**, where treatments often do not account for the **genetic diversity of the population.**
    - IGD enables **tailored healthcare solutions for India's demographic**, improving treatment effectiveness and outcomes.
- **Bioeconomy Growth:** IGD will **fuel India's growing bioeconomy,** which has surged from USD 10 billion in 2014 to over USD 130 billion in 2024.
- **Positioning India as a Global Leader**: India ranks 12th globally in biotech and 3rd in Asia-Pacific.
    - As the **largest vaccine producer with over 8,500 biotech startups** in 2023, India is set to lead the global bioeconomy.
    - IGD reduces dependence on foreign databases, positioning India as a hub for genomic innovation.
        - Under the **new BioE3 (Biotechnology for Economy, Environment and Employment) Policy,** this data will propel India to global leadership in biotechnology, advancing the nation's role in the **4th industrial revolution.**
- **Enhanced Genetic Tools: IGD** facilitates the creation of genomic tools and diagnostic tests specific to regional genetic variations, improving accuracy in healthcare.
- **Agricultural and Environmental Research:** Provides insights into genetic diversity that can improve crop breeding, disease resistance, and environmental sustainability.

What are India's Other Initiatives for Biotechnology Development?

- **BioE3 Policy**
- **National Biotechnology Development Strategy 2020-25**
- **National Biopharma Mission**
- **Biotech-KISAN Scheme**
- **Atal Jai Anusandhan Biotech Mission**
- **One Health Consortium**
- **Biotech Parks**
- **Biotechnology Industry Research Assistance Council (BIRAC)**
- **IndiGen Project:** Launched in 2019 by the **Council of Scientific and Industrial Research**

(CSIR), aims to map the genetic makeup of Indians to better understand disease susceptibility and drug responses.

Genomic Projects Across the World

- The **Human Genome Project,** an international collaboration funded by the US National Institutes of Health, completed the **world's first full human genome sequence in 2003.**
- The **EU's '1+ Million Genomes' (1+MG) initiative** seeks to provide secure access to genomic and clinical data across Europe, supporting research, health policy, and personalized healthcare to improve disease prevention.
- The **Earth BioGenome Project (EBP)** is a global initiative, aimed at sequencing and cataloging the **genomes of all known eukaryotic species on Earth**. This ambitious project is supported by India, China and the US.

Drishti Mains Question:

What is the significance of the Genome India Project (GIP) in the context of India's genetic diversity and healthcare system?

UPSC Civil Services Examination, Previous Year Questions (PYQs)

Q1. With reference to agriculture in India, how can the technique of 'genome sequencing', often seen in the news, be used in the immediate future? (2017)

1. Genome sequencing can be used to identify genetic markers for disease resistance and drought tolerance in various crop plants.
2. This technique helps in reducing the time required to develop new varieties of crop plants.
3. It can be used to decipher the host-pathogen relationships in crops.

Select the correct answer using the code given below:

(a) 1 only
(b) 2 and 3 only
(c) 1 and 3 only
(d) 1, 2 and 3

Ans: (d)