119TH CONGRESS
2D SESSION

# S. _____

To direct the Secretary of Energy to establish an extremophile whole genome sequencing initiative, and for other purposes.

---

IN THE SENATE OF THE UNITED STATES

---

Mr. ROUNDS (for himself and Mr. HEINRICH) introduced the following bill; which was read twice and referred to the Committee on _____

---

# A BILL

To direct the Secretary of Energy to establish an extremophile whole genome sequencing initiative, and for other purposes.

*Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,*

**SECTION 1. SHORT TITLE.**

This Act may be cited as the ''Homestake AI Act of 2026''.

**SEC. 2. EXTREMOPHILE WHOLE GENOME SEQUENCING INITIATIVE.**

(a) DEFINITIONS.—In this section:

(1) COVERED FACILITY.—The term ''covered facility'' means the long-term storage facility planned for under subsection (c)(1)(E).

(2) INITIATIVE.—The term ''initiative'' means the extremophile whole genome sequencing initiative established under subsection (b).

(3) SECRETARY.—The term ''Secretary'' means the Secretary of Energy, acting through the Under-secretary of Science.

(b) ESTABLISHMENT.—Not later than 90 days after the date of enactment of this Act, the Secretary shall establish an initiative, to be known as the ''extremophile whole genome sequencing initiative'', that shall—

(1) be consistent with the goals of the Genesis Mission announced in Executive Order 14363 (90 Fed. Reg. 55035; relating to launching the Genesis Mission); and

(2) facilitate the collection of new genomic data that will contribute to advances in scientific discovery and may support the development of new healthcare treatments.

(c) INITIATIVE.—

(1) IN GENERAL.—Under the initiative, the Secretary shall—

(A) collect, catalogue, and sequence the whole genomes and metabolites of samples from animals, plants, fungi, and microbes found in locations selected under paragraph (3);

(B) establish a database for the data collected under subparagraph (A), or use an existing database, such as the database of the National Center for Biotechnology Information, including sequence data, genome assemblies, peptides, metabolite data, annotations, and associated metadata, including genome and transcriptome data, metabolome data, and proteome data;

(C) collaborate with industry, academia, and other government agencies to collate existing relevant data for inclusion in the database established or used under subparagraph (B);

(D) make that database accessible to persons in the United States, subject to the condition that the Secretary shall, to the maximum extent practicable, prevent the sharing of such information with adversary nations or foreign persons; and

(E) not later than 1 year after the date of enactment of this Act, develop a plan for the

construction of a new long-term storage facility—

(i) for, at a minimum, the samples collected under subparagraph (A); and

(ii) in such a manner that the facility may continue to operate in the case of a catastrophic incident (as defined in section 602 of the Post-Katrina Emergency Management Reform Act of 2006 (6 U.S.C. 701)).

(2) THIRD PARTY COLLECTION.—The Secretary may enter into contracts with third party entities, including institutions of higher education or private sector entities, to collect samples under paragraph (1)(A).

(3) LOCATIONS FOR COLLECTION.—

(A) IN GENERAL.—Not later than 180 days after the date of enactment of this Act, the Secretary shall select not fewer than 5 locations from which to collect samples under paragraph (1)(A), at least 1 of which shall be a mine or former mine that—

(i) is deeper than 4,500 feet; and

(ii) houses an underground science facility.

(B) CONSIDERATIONS.—

(i) IN GENERAL.—The Secretary shall select locations under subparagraph (A)—

(I) that are unique or are exposed to unique properties, such as extreme heat or pressure;

(II) that are under risk of destruction before 2050 from natural or unnatural forces;

(III) based on the potential for finding useful organisms that will contribute to advances in scientific research and discovery; and

(IV) that may support the development of new healthcare treatments relating to cancer, including cancers affecting children and individuals with high mortality rates, diabetes, and dementia, including Alzheimer's disease.

(ii) PREFERENCE.—In selecting locations under subparagraph (A), the Secretary may give preference to a location on land owned or controlled by the Department of Energy (including National Laboratories).

(iii) RANGE OF EXTREME CONDITIONS.—In selecting locations under subparagraph (A), the Secretary shall ensure that the locations reflect a broad range of extreme conditions.

(4) LONG-TERM STORAGE FACILITY.—In selecting a location for the covered facility under the plan developed under paragraph (1)(E), the Secretary shall consider, among other factors determined relevant by the Secretary—

(A) colocation in a Department of Energy facility or a facility that receives significant funding from the Department of Energy;

(B) the benefits of storing the applicable materials underground in a remote area; and

(C) the ability of the location to withstand natural disasters, such as earthquakes, wildfires, or meteor strikes.

(5) DATA STANDARDS.—In carrying out the initiative, the Secretary shall—

(A) leverage existing and emerging data standards to ensure—

(i) high-quality data are obtained under the initiative; and

(ii) data collected under the initiative are considered usable and interoperable, with appropriate metadata to be included;

(B) ensure that data outputs under the initiative—

(i)(I) meet applicable Federal data standards for biological data; or

(II) if there are no applicable Federal data standards described in subclause (I), meet data standards for the initiative established by the Secretary, in consultation with the National Institute of Standards and Technology and other relevant agencies, based on findability, accessibility, interoperability, and reusability data principles and using industry best practices; and

(ii) are updated to comply with any applicable United States Geological Survey standards for biological data; and

(C) ensure that appropriate cybersecurity and access safeguards are implemented to protect the data generated or stored under the initiative, including by ensuring that data outputs under the initiative are updated as necessary to

meet applicable Federal cybersecurity standards for genomic data, including controls consistent with the National Institute of Standards and Technology special publications 800–53 (relating to Security and Privacy Controls for Information Systems and Organizations) and 800–111 (relating to Guide to Storage Encryption Technologies for End User Devices) or any successor framework.

(6) LONG-TERM STORAGE REQUIREMENTS.—The Secretary may consult with the Secretary of Agriculture and the Director of the National Museum of Natural History of the Smithsonian Institution with respect to storage capabilities for the management and long-term storage of samples collected under the initiative, subject to the following:

(A) Access to samples collected under the initiative may be managed by the respective institutional collection managers, with—

(i) the natural history collections managed by the Director of the National Museum of Natural History of the Smithsonian Institution; and

(ii) the germplasm repository and the covered facility managed by the Secretary of Agriculture.

(B) Before any samples are stored at the covered facility, the Secretary shall—

(i) evaluate the species to be sampled to identify which species need to be collected for long-term storage; and

(ii) from among the species identified under clause (i), identify which species should be considered to be high-priority species.

(C) Any samples identified for long-term storage under subparagraph (B)(i)—

(i) shall be collected and stored at the covered facility after the covered facility is operational; and

(ii) may be stored in an alternative location until that date.

(D) Any high-priority species identified under subparagraph (B)(ii) shall be collected for storage in each of—

(i) the natural history collections to be managed by the Director of the National

Museum of Natural History of the Smithsonian Institution;

(ii) the germplasm repository to be managed by the Secretary of Agriculture; and

(iii) the covered facility.

(E) Any physical samples collected under the initiative shall not be transferred, exported, or loaned outside the United States.

(F) Any physical storage, curation, and laboratory processing of samples collected under the initiative shall occur within facilities located in the United States.

(G) The Secretary may accept donated samples for storage from third parties or other countries.

(7) SEQUENCING AND DATA PROCESSING.—In carrying out the initiative, the Secretary shall—

(A) collaborate with the Secretary of Agriculture, the Director of the National Museum of Natural History of the Smithsonian Institution, and the heads of other applicable Federal agencies to sequence genomic information from samples collected under the initiative;

(B) develop and use standardized sequencing protocols to ensure consistency and comparability of data; and

(C) incorporate automated data processing and quality checks to ensure the reliability of the data.

(8) TERMINATION OF COLLECTION.—No collection of samples under paragraph (1)(A) may occur by the Secretary or a third party pursuant to a contract entered into under paragraph (2) after the date that is 5 years after the date of enactment of this Act.

(d) COORDINATION WITH OTHER ENTITIES.—

(1) IN GENERAL.—In carrying out the initiative, the Secretary may coordinate with—

(A) other offices within the Department of Energy;

(B) the Department of Agriculture;

(C) any other relevant agency that may benefit from utilizing the covered facility;

(D) the Smithsonian Institution, including the National Museum of Natural History;

(E) the Department of Health and Human Services, including the Centers for Disease Control and Prevention and the National Institutes

of Health, including the Childhood Cancer Data Initiative;

(F) academia;

(G) nonprofit organizations, including organizations focused on the collection of data for the discovery of treatments for cancer, diabetes, and dementia, including Alzheimer's disease;

(H) nonprofit organizations that support biodiversity conservation, including nonprofit organizations in allied countries described in subsection (e)(2); and

(I) private sector entities, including entities in the pharmaceutical, biotechnology, and artificial intelligence industries.

(2) PRIVATE SECTOR CONTRIBUTIONS.—There shall be established a foundation, which shall be an organization described in section 501(c)(3) of the Internal Revenue Code of 1986 and exempt from taxation under section 501(a) of that Code, to receive in-kind contributions from biotechnology and other private sector companies to assist in carrying out the initiative and supporting the operation of the covered facility.

(3) OTHER AGENCY RESOURCES.—The Secretary may use the services, personnel, and facilities of another agency if agreed to by that agency.

(e) WORKING GROUP.—

(1) IN GENERAL.—The Secretary shall establish a working group to coordinate with private and public sector initiatives in the United States and allied countries described in paragraph (2) collecting relevant genomic data to create alignment between scientific groups—

(A) to maximize beneficial research output; and

(B) to minimize duplicative spending of the initiative.

(2) ALLIED COUNTRIES DESCRIBED.—The allied countries referred to in paragraph (1) are—

(A) Australia;

(B) member countries of the European Union;

(C) Japan;

(D) South Korea;

(E) New Zealand; and

(F) the United Kingdom.

(f) AUTHORIZATION OF APPROPRIATIONS.—There are authorized to be appropriated to the Secretary such sums as are necessary to carry out this section.