NATURE|Vol 464|1 April 2010 HUMAN GENOME AT TEN **NEWS FEATURE**

HUMAN GENOME

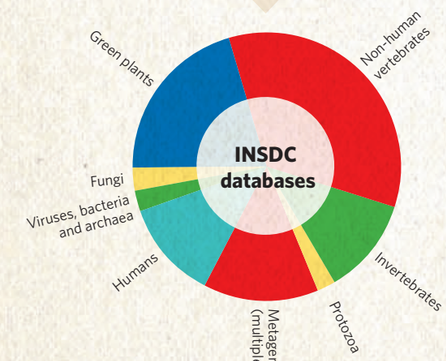
THE SEQUENCE EXPLOSION

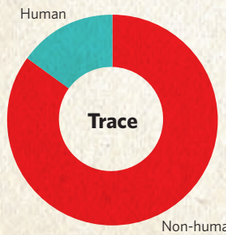
At the time of the announcement of the first drafts of the human genome in 2000, there were 8 billion base pairs of sequence in the three main databases for 'finished' sequence: GenBank, run by the US National Center for Biotechnology Information; the DNA Databank of Japan; and the European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Database. The databases share their data regularly as part of the International Nucleotide Sequence Database Collaboration (INSDC). In the subsequent first post-genome decade, they have added another 270 billion bases to the collection of finished sequence, doubling the size of the database roughly every 18 months. But this number is dwarfed by the amount of raw sequence that has been created and stored by researchers around the world in the Trace archive and Sequence Read Archive (SRA).

See Editorial, page 649, and human genome special at www.nature.com/humangenome

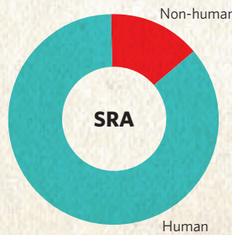
DNA SEQUENCES BY TAXONOMY

International Nucleotide Sequence Database Collaboration: The main repositories of 'finished' sequence span a wide range of organisms, representing the many priorities of scientists worldwide.

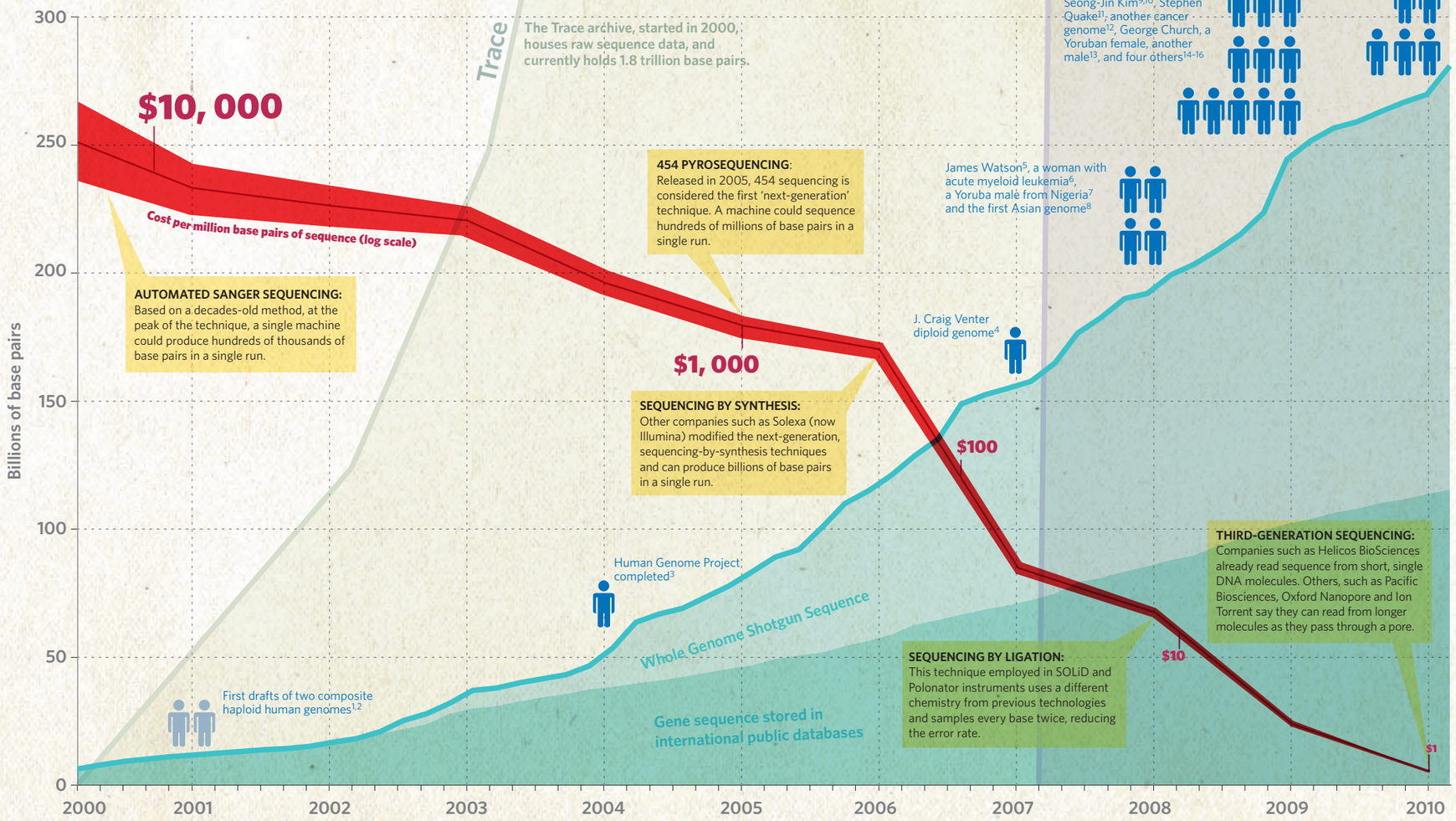




Trace Archive: Developed to house the raw output of high-throughput sequencers built in the late 1990s, the trace archive spans a wide range of taxa.



Sequence Read Archive: Houses raw data from next-generation sequencers. Dominated by human sequence, including multiple coverage for more than 170 people.



SOURCE: NCBI; GRAPHICS BY N. SPENCER & W. FERNANDES

HOW MANY HUMAN GENOMES?

The graphic shows all published, fully sequenced human genomes since 2000, including nine from the first quarter of 2010. Some are resequencing efforts on the same person and the list does not include unpublished completed genomes.

1. Venter, J. C. *et al. Science* **291,** 1304–1351 (2001).
2. International Human Genome Sequencing Consortium *Nature* **409,** 860–921 (2001).
3. International Human Genome Sequencing Consortium *Nature* **431,** 931–945 (2004).
4. Levy, S. *et al. PLoS Biol.* **5,** e254 (2007).
5. Wheeler, D. A. *et al. Nature* **452,** 872–876 (2008).
6. Ley, T. J. *et al. Nature* **456,** 66–72 (2008).
7. Bentley, D. R. *et al. Nature* **456,** 53–59 (2008).
8. Wang, J. *et al. Nature* **456,** 60–65 (2008).
9. Ahn, S.-M. *et al. Genome Res.* **19,** 1622–1629 (2009).
10. Kim, J.-I. *et al. Nature* **460,** 1011–1015 (2009).
11. Pushkarev, D., Neff, N. F. & Quake, S. R. *Nature Biotechnol.* **27,** 847–850 (2009).
12. Mardis, E. R. *et al. N. Engl. J. Med.* **10,** 1058–1066 (2009).
13. Drmanac, R. *et al. Science* **327,** 78–81 (2009).
14. McKernan, K. J. *et al. Genome Res.* **19,** 1527–1541 (2009).
15. Pleasance, E. D. *et al. Nature* **463,** 191–196 (2010).
16. Pleasance, E. D. *et al. Nature* **463,** 184–190 (2010).
17. Clark, M. J. *et al. PLoS Genet.* **6,** e1000832 (2010).
18. Rasmussen, M. *et al. Nature* **463,** 757–762 (2010).
19. Schuster, S. C. *et al. Nature* **463,** 943–947 (2010).
20. Lupski, J. R. *et al. N. Engl. J. Med.* doi:10.1056/NEJMoa0908094 (2010).
21. Roach, J. C. *et al. Science* doi:10.1126/science.1186802 (2010).

Page size by comparison