

The Geography of the Genome

DNA SPACE: Welcome to the Biotechnomy

By Rodrigo Martinez, Juan Enriquez, and Jonathan West

The most powerful code is no longer a string of 1s and 0s. It is As, Ts, Cs, and Gs: adenine, thymine, cytosine, and guanine. During the past 10 years, there has been a boom in the generation and storage of raw genetic data. A good lab or genetics company creates the equivalent of more than eight times the printed collection of the Library of Congress every single month. And the rate is soaring. Compare the mounting number of As, Ts, Cs, and Gs in GenBank, the National Institutes of Health's sequence database, against a graph of Moore's law, for instance, and suddenly Moore goes flat.

In the biotech community, the focus is now on mining genetic and proteomic data. Unlike oil or electricity, most of this data is neither scarce nor expensive. Anyone can access it, anyone can use it. As a result, researchers expected genetic information to become a global resource, shared equally. In fact, this hasn't happened.

It turns out a new world hierarchy is developing, one that separates those nations and regions that are bioliterate from those that are bio-illiterate. This is the world of DNA space, populated by a self-selecting few who have chosen to participate in the new technology revolution. The price of admission: the ability to produce, read, or translate DNA. This means that even as biodata begins to drive industries from agribusiness to computing, cosmetics to chemical manufacturing, few nations have the skills required to develop, access, and use it.

To assess that conclusion, the Harvard Business School Life Sciences Project mapped which countries and domains massively accessed (via FTP) the three largest public biodatabases in the world in September, October, and November 2000 and 2001. During these six months, countries checked out a lot of data: 43 terabytes, approximately seven times more than was downloaded over a comparable period from the Library of Congress.

Yet it turns out that few countries understand or read large-scale

PLUS

[The New World](#)

[Euro Space](#)

[Nano Space](#)

[Space Space](#)

[Relationship Space](#)

[Dump Space](#)

[Atlas Space](#)

[Voice Space](#)

[Office Space](#)

[Home Space](#)

[Bush Space](#)

[Protest Space](#)

[Boom Space](#)

[Body Space](#)

[Research Space](#)

[Tight Space](#)

[Art Space](#)

[Sex Space](#)

biodata. Nearly all - 92 percent - of the data was downloaded by users in 10 countries, half of whom live in the US. Europe, which has similar population, education, and income figures, accessed just 22 percent of the data. No country in Africa, Latin America, the Middle East, or Asia (except Japan) downloaded 1 percent or more of this free information.

[Border Space](#)[Crowd Space](#)[Future Space](#)

And who, in particular, is using the information? Within the US, we expected the leading consumers of large new data sets to be primarily research scientists, mostly at universities. Again we were surprised. Users from .com domains downloaded about half of all US data versus 38 percent going to universities (.edu).

[Secure Space](#)[Color Space](#)[Blog Space](#)[Waning Space](#)[Robo Space](#)[DNA Space](#)[Ad Space](#)[Golf Space](#)

Biogaps are growing not just among countries, but within them. The top five biotech patent-producing states account for 57 percent of all such patents in the US. And, looking at where biotech companies are headquartered, there's a concentration of them in very few zip codes: 92121 (San Diego County, California), 94080 (South San Francisco, California), 20850 (Rockville, Maryland), and 02139 (Cambridge, Massachusetts).

[Limbo Space](#)

A single organization was responsible for 93 percent of all Canadian downloads from the European database. The least concentration occurred in the UK, where the biggest user accounted for 15 percent of downloads, and the top three for a modest 34 percent.

[Public Space](#)

Meanwhile, Japan is in an info deficit, importing more than it exports. As the second-largest global users of biodata, Japan-based organizations are huge accessors of US and European databases. Yet Japan's own version, DDBJ, is underutilized, even by Japanese. Mostly universities, not businesses, are doing the downloading in Japan. This is odd in a nation where business funds more than 70 percent of all research. Downloads from .co domains (equivalent to US .coms) accounted for only 0.5 percent of total files. Educational institutions made up 93 percent of the 2001 downloads.

Columbus' first squiggly outline of the Americas couldn't foresee the rise of the Spanish empire or the spread of diseases across the hemisphere. Likewise, this map of biodata flows doesn't tell us who will be successful in the life sciences revolution and who won't. As information leaves wet labs and silicon-research facilities, we will no doubt see the world, and life, from a different perspective. This map is just the first draft.

Rodrigo Martinez, Juan Enriquez, and Jonathan West are founding members of the Harvard Business School Life Sciences Project.

[Copyright](#) © 1993-2004 The Condé Nast Publications Inc. All rights reserved.

[Copyright](#) © 1994-2003 Wired Digital, Inc. All rights reserved.