

his colleagues to come up with a way to analyze this previously impenetrable DNA.

As described on page 641, Eichler and his colleagues have developed a technique for counting the number of copies of a gene in any duplicated region. The genes can vary in copy number among people. The number can affect how much of that gene's protein is produced, and consequently, the function of that protein.

Eichler's team has also come up with a way of distinguishing near-identical cop-

ies. Over time, copies tend to develop slight sequence differences that could also affect how that gene—or its protein product—works. Eichler's team has cataloged these telltale variant bases for about 70% of the duplicated genes. "It's opened up a whole new area of genetic diversity that we have not been able to tap previously," says Eichler.

Surprisingly, analysis of the 1000 Genomes data showed quite large differences in the copy number of certain genes between the African, European, and Asian populations, Eichler

reports. "Humans are more different than we would have ever thought," says Eichler.

"Once all this variation is revealed, it changes the way you can think about [doing] genetics," notes 1000 Genome Project co-leader Richard Durbin of the Sanger Institute. For most of the history of genetics, researchers have been fishing out variation without knowing what was there. Now, "we are right on the cusp where we do genetics in the light and [see] exactly what it is that we are studying."

—ELIZABETH PENNISI

## IRAQ WAR

# Leaked Documents Provide Bonanza for Researchers

The Pentagon is fuming after last week's release of a huge cache of classified Iraq War data by the organization WikiLeaks. But researchers struggling to build an accurate picture of the death toll in post-invasion Iraq are thrilled. "It is hard to overstate the significance of this development for the conflict field," says Michael Spagat, an economist at Royal Holloway, University of London, U.K.

Within the nearly 400,000 leaked documents is a stream of raw data called SIGACTS—for Significant Activities—that chronicles the casualties directly observed by U.S. soldiers in Iraq. In late summer, WikiLeaks passed a copy of these data to Iraq Body Count (IBC), a London-based organization that has tallied the war's death toll using media reports of casualties. Their numbers do not include insurgents or soldiers. And because not every violent death is reported in the media, IBC's numbers are known to be an underestimate of the true number of war dead. But how much higher the true number is has been a source of intense debate, with surveys of Iraqi households yielding a wide spread of casualty estimates (*Science*, 20 October 2006, p. 396). As *Science* went to press, the IBC toll for Iraqi civilians stood at 98,585 to 107,594 violent deaths.

According to the IBC analysis of the leaked SIGACTS data, published online on 25 October, more than 109,000 violent deaths in Iraq were logged by the U.S. military between January 2004 and December 2009. Of these, over 79,000 were civilian deaths comparable to those logged by IBC, which recorded about 91,000 over the same period. By extrapolating from a sample of the data, IBC estimates that at least 27,000 civilian deaths went unrecorded by the U.S. military, while the military observed 15,000 comparable deaths that the media missed. Most of these unreported deaths were from small inci-



**Buried data.** WikiLeaks' Julian Assange at a briefing on the release of classified Iraq war records.

dents of violence, with between one and three casualties. This confirms a widely assumed bias in media reporting in favor of larger incidents, such as suicide bombings. "But with such a huge overlap, it does not seem very likely that there are a large number of civilian deaths missed by both sources," says Spagat, who helped IBC with its analysis.

Taking the WikiLeaks data into account, IBC now estimates that at least 150,000 have died violently during the war, 80% of them civilians. That falls within the range produced by an Iraq household survey conducted by the World Health Organization—and further erodes the credibility of a 2006 study published in *The Lancet* that estimated over 600,000 violent deaths for the first 3 years of the war (*Science*, 18 January 2008, p. 273).

The leaked data are sure to keep researchers busy for months to come. Besides the number of casualties, the SIGACTS release includes geographic locations of the vio-

lence and other information that has not been available until now. But there could be serious challenges to those hoping to publish an analysis, says Gary King, director of the Institute for Quantitative Social Science at Harvard University. "I have had a couple of students asking [Harvard] for permission to use the previous WikiLeaks data release, and last I heard they still weren't allowed to touch it." But others are more optimistic. "As long as the data is stripped of information that could be used to identify anyone, it shouldn't be a problem," says Christian Davenport, a political scientist at the University of Notre Dame in Indiana who studies conflict mortality. "What's important is that people appreciate the complexity." For example, he says, "we don't know exactly how these data were gathered." It represents the results of an experiment, "but we don't know the methods."

—JOHN BOHANNON