



Three researchers in Jia Wei's lab with the surname Wang, Xiao-yan, Xiao-rong and Xiao-xue (left to right, with Chinese names above), all publish in English as X. Wang.

## Identity crisis

Chinese authors are publishing more and more papers, but are they receiving due credit and recognition for their work? Not if their names get confused along the way. **Jane Qiu** reports.

Jia Wei, associate dean at the pharmacy school of Shanghai Jiao Tong University can remember hundreds of metabolic pathways by heart, but he gets confused by his graduate students' publications. Three of his students — Wang Xiao-yan, Wang Xiao-rong and Wang Xiao-xue (pictured above with Jia) have completely different two-character given names in Chinese, but all publish under the abbreviated name X. Wang. "I really have a hard time sorting out who has published what," Jia sighs.

A similar confusion could arise if John Roberts and Jane Roberts worked in the same lab and both published as J. Roberts. But name recognition in China is compounded by the challenges of transliterating Chinese characters for English-language publications, and by overuse of a few common surnames by the growing population. Estimates by China's Ministry of Public Security suggest that more than 1.1 billion people — around 85% of China's population — share just 129 surnames. Problems with abbreviations, ordering of given names and surnames and inconsistent journal practices heighten the confusion.

When publishing in English-language

journals, Chinese researchers adopt a phonetic version of their names, converted through the pinyin romanization system, which uses the Latin alphabet to represent sounds from Chinese. This approach, however, is not bidirectionally unique. There are two Chinese surnames that can be 'spelt' as Wang, for instance. And the problem is compounded by the sheer number of Chinese researchers who have not just the same surname, but also the same initial. Searching the bio-medical-literature database PubMed, curated by the US National Library of Medicine, for articles published by 'Wang X' results in 8,904 entries, and this number rises almost daily.

This issue is not unique to the Chinese. "Japanese and Korean names have the same problem when published in English," says Masao Ito, president of the Human Frontier Science Program based in Strasbourg, France, which promotes international research in the life sciences. Not surprisingly, researchers and editors using search engines and publication databases find it difficult to identify Asian authors. "As a result,

Asian researchers are less likely to be invited to participate in collaborative projects or to become reviewers," says Ito.

Publishers make things worse by having varying rules for Asian names. For example, journals differ in how they abbreviate polysyllabic Asian names. If journals abbreviated all the Chinese characters of a given name (Xiao-rong becoming X. R. and Xiao-xue becoming X. X., rather than shortening them to just X.), Jia says that it would help to distinguish between researchers' publications. "This is very problematic when we appraise researchers' performance or during head-hunting," he says.

Similarly, some journals insist on listing given names first and surname second, whereas others allow authors to present their names according to the tradition in their own countries. Take, for instance, two researchers previously working on nanocarbon technology at Tsinghua University in Beijing, Yang Wei and Wei Yang (see 'Character confusion'). According to Yang Wei, now at Zhejiang University in Hangzhou, not only did several researchers at Tsinghua

**"I really have a hard time sorting out who has published what."**

— Jia Wei

have exactly the same name as his, but he shares the same initial with several other Yang's working on nanocarbons elsewhere. "You would be lucky to be able to locate the researcher you are looking for," he says.

The problem is sufficiently widespread that some researchers have taken advantage of the ambiguity. Surgeon Liu Hui, who padded his CV with publications by another researcher who shared his surname and initial, rose to become an assistant dean at the prestigious Tsinghua University. But the discrepancies were noticed and he was dismissed by the university in March 2006.

And if Asians can't distinguish between researchers from their own country, it's much more challenging for Westerners, says Gene Sprouse, editor-in-chief of the research journals for the American Physical Society. "When I asked my editors why we have so few Asian reviewers, they said that it's because so many Asian researchers have the same surname and initial that they have difficulties in pinpointing the appropriate ones," Sprouse says. And the problem will only get worse as Asian authors publish more papers, he adds.

Publications from China, Japan and South Korea have increased rapidly in recent years, and by 2006 made up one-fifth of the scientific literature indexed by Thomson Scientific's Science Citation Index (SCI) — roughly two-thirds of the amount from the United States (L. Leydesdorff and C. Wagner *Scientometrics*; in the press). Publications by authors in mainland China indexed by the SCI are growing particularly fast, from 2.3% in 1996 to 8.4% in 2006.

### More than words

To address this trend, the American Physical Society has taken the unusual step of offering its authors the option to list their names in Chinese, Japanese or Korean characters, in addition to the transliterated English version. "This is not just a publishing issue," says Sprouse. "A person's name is important. Our initiative is a statement that we respect our Asian colleagues and welcome their submissions to our journals." The society may extend the offer to other languages, such as Arabic.

Although many welcome the society's move, some doubt whether other publishers will rush to adopt the strategy. To print in several foreign alphabets, typesetters would need to incorporate fonts and codes for every one of them, raising costs. Others question how helpful the approach will be on its own. "It would be truly useful only if it's taken up by publication databases so Asian authors could be searched in their own language," says Ito.

These concerns do not yet seem to be troubling literature databases such as PubMed. James Marcetich, head of the index section at the National Library of Medicine, says that he is not aware of problems associated

## Character confusion

Apart from unhelpful abbreviations (see image opposite), there are two other ways in which conversion of Chinese names into English can cause problems.



Li Yan

Li Yan

**Phonetic confusion** Two researchers with different Chinese names both have their names transliterated as Li Yan in English.



Wei Yang

Yang Wei

**Mixed-up order** Researchers named Yang Wei and Wei Yang can be confused if journals have different rules for the ordering of given names and surnames.

with searching citations with Asian authors. In many cases users can narrow down their search by typing in keywords or the author's affiliation. However, many researchers change affiliations every few years or work in multiple fields. And the narrowing-down process can be daunting and time-consuming.

"It's all about efficiency," says Sun Xiao-peng, a manager at the Beijing office of the Dutch science and technical publisher Elsevier. "It's a matter of whether an author can be located in 15 minutes or 6 hours." He doubts whether searching in native languages would help because it's extremely difficult for researchers to search names in a language unfamiliar to them. He says that Elsevier has tools — including its Scopus author identifier — that provide an alternative way to identify both Asian and Western authors with common surnames.

Launched in 2006, the author identifier assigns a unique number to the roughly 20 million authors who have published articles in the 15,000 journals covered by the Scopus database. An algorithm distinguishes those with similar or identical names on the basis of their affiliations, publication history, subject areas and co-authors. Scopus claims to have achieved 99% certainty for 95% of its records, and designates a webpage to each

author so that they have an overview of their personal data and can make corrections. "This is akin to a researcher's passport, which follows them wherever they go," says Sun.

Elsevier is not the only company offering such commercial products. Last month, Thomson Scientific in Philadelphia unveiled its own ResearcherID, which allows researchers to create stable personal identifiers to manage their citation metrics. The software allocates a number during a one-time registration. "ResearcherID resolves any ambiguity surrounding published works and provides a safe space for virtual collaboration," says Jim Pringle, Thomson's vice-president of product development. The service currently has some 3,500 invited users.

### United front

Librarian Susan Fingerma at Johns Hopkins University's Applied Physics Laboratory in Laurel, Maryland, thinks that Elsevier's claim of accuracy in identifying authors "is probably overblown". In any case, she argues, these approaches can provide only part of the solution because the problem of author identification lies farther upstream in publishing — with journals having inconsistent rules for Asian names. She thinks that publishers should get together to agree on a uniform approach to author names. "Only then could author identification systems be truly useful," she says.

The products currently in the market, however, fall short of providing an identifier that works across all databases. One solution may come from CrossRef — a coalition of 2,046 scholarly publishers, including Elsevier and Nature Publishing Group — that wants to introduce a Contributor ID, similar to its digital object identifiers (DOIs) for electronic content. "CrossRef has been considering the idea of a universal author DOI registration for sometime," says Geoffrey Bilder, CrossRef's director of strategic initiatives. Bilder says that both algorithm and user-generated systems are prone to errors and what CrossRef can offer is authentication, which is essential if an identifier is to be used for professional purposes.

Later on this year, CrossRef will trial a prototype system that will invite contributors — including authors of books and academic papers — to register their material. The publications listed by the authors will then be authenticated through linking to publishers' websites.

"A universal author-identification system could help to clarify the confusing situation," says Wang Xiao-yan. There are some 93 million people in China who share the surname Wang, and so the chances of her working alongside, or in the same field, as another X. Wang are high. She worries that misidentification issues could prevent her from competing with her Western colleagues on an equal footing. "This could be a stumbling block to my career," she says. ■

Jane Qiu writes for *Nature* from Beijing.

**"You would be lucky to be able to locate the researcher you are looking for."**

— Yang Wei