Increasing global participation in genetics research through DNA barcoding
Increasing global participation in genetics research through DNA barcoding

By

Sarah J. Adamowicz¹² & Dirk Steinke¹

¹ Biodiversity Institute of Ontario, University of Guelph, 50 Stone Road East, Guelph, ON, N1G 2W1, Canada.

² Department of Integrative Biology, University of Guelph, 50 Stone Road East, Guelph, ON, N1G 2W1, Canada.

Email addresses: sadamowi@uoguelph.ca; dsteinke@uoguelph.ca
Abstract

DNA barcoding—the sequencing of short, standardized DNA regions for specimen identification and species discovery—has promised to facilitate rapid access to biodiversity knowledge by diverse users. Here, we advance our opinion that increased global participation in genetics research is beneficial, both to scientists and for science, and explore the premise that DNA barcoding can help to democratize participation in genetics research. We examine publication patterns (2003-2014) in the DNA barcoding literature and compare trends with those in the broader, related domain of genomics. While genomics is the older and much larger field, the number of nations contributing to the published literature is similar between disciplines. Meanwhile, DNA barcoding exhibits a higher pace of growth in the number of publications as well as greater evenness among nations in their proportional contribution to total authorships. This exploration revealed DNA barcoding to be a highly international discipline, with growing participation by researchers in especially biodiverse nations. We briefly consider several of the challenges that may hinder further participation in genetics research, including access to training and molecular facilities as well as policy relating to the movement of genetic resources.

Keywords: DNA barcoding, International Barcode of Life, genomics, research trends, collaboration
The benefits of increasing global participation in genetics research

The invitation to contribute an opinion piece to the special issue entitled “Genomics: the Power and the Promise” inspired us to reflect upon several different axes of benefits. Here, we focus upon the power of genomics and related disciplines to enable increased participation in research and to link together members of the international research community. We advance our opinion that participation in genetics research—of all types—is beneficial. We first briefly consider some of the benefits that arise from being a participant in the research endeavour, on top of those gains that would be incurred anyhow as long as someone or anyone is conducting such research (e.g. better medicines administered; improved environmental health; etc.).

We suggest that there are myriad benefits to individuals, nations, and humanity from participating in genetics research. Individual scientists benefit from leading or collaborating in research that is of general interest to other members of the global research community. Publishing in international journals, which can be readily accessed via the most commonly-used databases of scientific literature, helps individual scientists to attain citations, apply for funding, and advance their careers in various other ways, such as via international networking and further collaboration. Nations that are home to scientists connected to the global research community benefit as well, due to increased access to cutting-edge knowledge and methodologies, which can be directed towards addressing national priorities through granting programs and other means. When opportunities to participate in science are available broadly, more individuals have the possibility to realize their full potential for creativity, discovery, and productivity, advancing global science together with improving their own human experience. As
well, participation by individuals inhabiting different regions enables scientific advances simply not possible through intensive investigation of just a small part of the planet. As just a few of many possible examples, including more of Earth’s biodiversity in research enables new studies of evolutionary history, diversification, dispersal, and macroecology. Biomes such as the world’s oceans naturally connect multiple nations and can only be understood through international collaboration.

While several of the above advantages are not unique to the disciplines of genetics and genomics, we think that the benefits—to individuals, societies, and science at large—from increasing the participant base are particularly strong in these disciplines. For example, species do not respect national boundaries; therefore, many objectives relating to biological monitoring (such as detecting invasive and quarantine species) require global research efforts to create comprehensive species and sequence databases to enable rapid identification. Moreover, sequenced DNA is a type of data that can be easily shared via the web and analyzed and re-analyzed for many purposes. Finally, as all nations harbour some degree of unique genetic diversity, within humans and especially within and among other species, the existence of active scientists from a diversity of nations is important for national and international discussions of policies on genetic resources.

While each of these proposed benefits could be explored in detail in its own right, we are assuming here that, on balance, we are correct that an increase in participation in genetics research is a “good” thing. In the balance of this contribution, we explore the idea that one
particular research area, DNA barcoding, is an excellent conduit for facilitating an increase in such participation.

The promises of DNA barcoding

DNA barcoding is the approach of using short, standard DNA sequences for the identification of specimens to species as well as for the discovery of new species (Hebert et al. 2003a; Collins and Cruickshank 2013). While the term “DNA barcodes” was previously introduced in the literature (Arnot et al. 1993; Floyd et al. 2002), 2003 (Hebert et al. 2003a, 2003b) is generally considered the year of origin of DNA barcoding as a standardized, large-scale proposed initiative. The power of the barcode approach—enabled by its simplicity and by standardization—is that barcoding efforts around the world will contribute to our collective ability to recognize species for addressing diverse research questions and applications.

Research trends (Adamowicz 2015) indicate increasing usage of DNA barcoding not only in research but also for socio-economically important applications, including the development of quarantine applications and agricultural pest identification systems (e.g. Hodgetts et al. 2015; Wang et al. 2015); enforcement of endangered species laws (Holmes et al. 2009; Dalton and Kotze 2011; Gonçalves et al. 2015); marketplace product authentication systems for seafoods (Hanner et al. 2011; Cawthorn et al. 2015), herbal medicines (e.g. Newmaster et al. 2013; Seethapathy et al. 2015), and timber trees (Dev et al. 2014; Nithaniyal et al. 2014); and studies of pathogen/vector/host species associations (Gariepy et al. 2012; Brugman et al. 2015).

DNA barcoding additionally promises to be a conduit for global engagement in genetics and biodiversity research. Due to the now well-established barcoding pipelines for many
taxonomic groups (e.g. Bucklin et al. 2011; Hollingsworth et al. 2009; Hebert et al. 2003a; Pawlowski et al. 2012; Schoch et al. 2012), the barriers to entry for new researchers are modest. Moreover, centralized sequencing facilities can facilitate barcoding-enabled research by those who do not have a genetics laboratory in their home institution or by those wishing to participate in large-scale barcoding projects. Furthermore, the online data repository and workbench tools available through the Barcode of Life Data Systems (BOLD; Ratnasingham and Hebert 2007) provide online data sharing and collaboration. However, is this potential for international participation in research being realized?

We explore whether there is emerging evidence that DNA barcoding is indeed associated with increasing international participation in genetics research. We summarize publication trends in the DNA barcoding literature and compare these with patterns in another major contemporary scientific field of study, genomics. DNA barcoding has been referred to as horizontal genomics (Waterton et al. 2013), emphasizing the fact that it largely relies on a single gene or—in the case of plants—very few genes, while genomics involves entire genomes. More recently, these clear boundaries have been somewhat blurred, as both use similar technology and terminology. For the sake of our study, we maintain the distinction of the one gene, many taxa approach (DNA barcoding) and the many genes, one taxon concept (genomics). Due to its recent emergence and lower barriers to entry relative to genomics, we predict that the discipline of DNA barcoding will be characterized by more rapid growth in publications, a higher diversity of nations represented among authors of scientific publications, and more authorships from researchers based in developing nations (Janzen and Hallwachs, 2011). We will discuss the
findings in the context of emerging research directions and possibilities for further development.

A synthesis of publication statistics: barcoding compared with genomics

Methodology

Using the ISI Web of Knowledge publications databases, two literature searches were performed in July 2015 for papers published between 2003 and 2014. The search terms were “DNA barcod*”, which yielded 4,228 total papers, and “*genomics”, which yielded 132,659 total papers from this date range. The asterisk (*) wildcard character was used to search for any string of characters. Redundant entries were excluded and are not included in the counts above. The proportion of total publications from the defined date range for each discipline was graphed per year, to reveal trends in the growth rate of publications (Fig 1). The citations were downloaded, and the country of the home institution for each author determined. Each author on every paper was counted equally here as a single unique paper/author data point, regardless of the total number of authors per paper. The number of countries represented among the publishing authors was plotted by year to show trends over time (Fig 2), and evenness in the distribution of countries among total authorships was also investigated (Fig 3 and Fig 4). Four salient patterns were apparent among the publication results.

Pattern #1: Higher growth rate of DNA barcoding publications
While genomics is the older discipline and considerably larger in terms of total publications from 2003-2015, the growth rate in DNA barcoding publications is higher (Fig 1). After its inception in 2003, the growth curve for DNA barcoding publications approximated an exponential curve shape until ca. 2012, followed by a modest leveling off in the rate of growth. 19% of the total barcoding papers from 12 years (2003-14) were published in 2014, while this value was 11% for the genomics literature. Growth in the barcoding community is also apparent when considering growing participation in the biennial International Barcode of Life Conference series (Adamowicz 2015). Extrapolating from these trends, one might predict that research in DNA barcoding will continue to grow over the coming years.

Pattern #2: Growth in the number of publishing nations

Both disciplines show growth in the number of nations publishing scientific articles from 2003-2014, with DNA barcoding starting from a lower number of nations, as expected, and exhibiting rapid growth (Fig. 2). Interestingly, in 2014 there were 94 nations represented among the publishing authors in both disciplines. Moreover, over the total time period examined, the number of publishing nations was similar between disciplines, with 113 nations represented among the authors of barcoding publications and 117 nations represented in genomics. The lists of nations overlap considerably, as there are 123 total nations represented considering both disciplines together. It is not clear whether the participation in scientific publishing, in terms of national representation, will continue to grow for DNA barcoding. It is possible that this growth will level off to the same approximate number of participating nations that
genomics has witnessed since ca. 2009 (Fig. 2). We might predict that some further growth is
feasible within barcoding, due to the lower barriers to entry. Molecular methods are
standardized and well developed for many taxa, which can be implemented in various settings,
and we suggest it is also simpler to enter biological samples into the processing pipelines for
medium-throughput (e.g. Martínez-Arce and Elías-Gutiérrez, 2015) and high-throughput (e.g.
Biodiversity Institute of Ontario, University of Guelph) barcoding facilities than for genomics.

**Pattern #3: Greater evenness of authorships among nations for DNA barcoding**

While the total number of participating nations is similar for DNA barcoding and
genomics, there are substantially different patterns when considering evenness. In genomics,
45% of all authorships are by authors with an institutional address based in the USA. Moreover,
just 6 nations together make up more than 75% of all authorships. By contrast, for DNA
barcoding, the efforts of authors from 15 nations must be included to surpass the 75% mark.
Interestingly, the lists of top 25 publishing nations in both disciplines include several especially
biodiverse nations (e.g. China, India, Brazil, and Mexico). In each case, these biodiverse nations
are more strongly represented among authors of DNA barcoding publications. Canada,
recognized as the birthplace of the large-scale DNA barcoding movement (Hebert et al. 2003a),
is more prominent among barcoding (9%) than genomics (6%) authorships, but overall Canada
represents a modest fraction of total barcoding authorships. This result likely reflects the
international as well as collaborative nature of DNA barcoding research.

Despite the greater evenness for DNA barcoding, it was interesting to observe that the
top six nations in each discipline were the same, with Japan 7th for genomics and Australia 7th in
barcoding. We suggest that this reflects general levels of participation and publishing in science, rather than a direct association between these lines of research. In particular, at least in the majority of cases, we do not think that previously existing genomics facilities in these countries are being extensively used for DNA barcoding as well. It is a very different workflow to generate a single sequence from a million different specimens, as compared with generating millions of sequences from a single sample. Unique aspects of high-throughput DNA barcoding include efficient “front-end” specimen processing, enabling integration with the molecular pipeline, as well as high-throughput DNA extraction of many unique specimens (Ivanova et al. 2006; Borisenko et al. 2009; Hebert et al. 2013). By contrast, DNA barcoding facilities and several national barcoding networks are increasingly adding tools that are commonly used in genomics, including next-generation DNA sequencers, for whole-community metabarcoding and, increasingly, for mitogenomics, chloroplast genomics, and metagenomics as well. For example, 20% of the abstracts of the 6th international Barcode of Life Conference either employed or related to developing next-generation sequencing (Adamowicz 2015).

**Pattern #4: Continental leadership outside North America in DNA barcoding**

Interesting regional patterns are also apparent upon considering authorships grouped by continent, per discipline (Fig. 4). North American leadership among all authorships in genomics is overwhelming (particularly the USA; Fig. 3). In striking contrast, all other continents display greater proportional participation in publications relating to DNA barcoding than genomics.
Finer examination of research patterns also reveals varying regional research priorities. China has emerged as a leader in plant DNA barcoding (Bubela et al. 2015), while India and South Africa also display substantial research activities in this domain, with all three countries intensively using DNA barcoding to investigate the diversity and authenticity of medicinal plants (e.g. Lekganyane et al. 2015; Ratsoma et al. 2015; Seethapathy et al. 2015; Yao et al. 2015; see also Genome 2015 for many abstracts on this theme from the 6th International Barcode of Life Conference). Another example is the current leadership of South America in bird barcoding, for example using barcodes to investigate bird diversity and geographic patterns of diversification (e.g. Lavinia et al. 2015a), to quantify rates of molecular evolution (e.g. Lavinia et al. 2015b), and to reveal illegal trade in live parrots (Gonçalves et al. 2015).

**Growth of the International Barcode of Life community**

Our brief synthesis of publication statistics largely supported our original thesis that international participation in DNA barcoding research has been rapidly growing. The number of nations involved now equals participation in genomics research, but with substantially greater proportional participation of authors residing outside the USA. More broadly, multiple middle-income nations (per capita) have developed mature DNA barcoding programs and represent a substantial proportion of all authorships in DNA barcoding. We did not examine in detail the reasons for this growth, but one may presume that the barriers to entry into single- or two-marker genetic work are significantly lower than for genomics. Moreover, researchers may participate in barcoding activities in collaboration with centralized sequencing or barcoding
facilities. For example, 1,064 of 4,228 total barcoding publications (25% of the total) acknowledge the Biodiversity Institute of Ontario or the Canadian Centre for DNA Barcoding, University of Guelph for providing sequencing support. This percentage vastly exceeds the proportion of Canadian authorships of barcoding publications (9%), and thus the centralized facility is facilitating global collaboration as well as access to high-throughput DNA infrastructure. Regionally important centralized molecular facilities are also emerging in other countries, such as Mexico (Martínez-Arce and Elías-Gutiérrez 2015).

Another factor to consider with respect to differences among disciplines is that biodiversity, by its very nature, is global. There are endemic assemblages of species on all continents and even within nations. Therefore, to be globally useful for both research and management (e.g. of international trade), DNA barcode reference libraries should include all corners of the world. Such libraries as well as international collaboration have been actively cultivated by participants in the International Barcode of Life initiative (www.ibol.org) and through international conferences (see *PLoS One* 2011 for the special collection of articles arising from the Third International Barcode of Life Conference held in Mexico City, Mexico; *Genome* 2015 for the special issue of abstracts for the 6th International Barcode of Life Conference held in Guelph, Canada). Moreover, the DNA barcoding community has been keener to promote wide access to the research results, as 17.6% of the total publications are available in open-access journals, compared to 8.4% for genomics. Thus, some of the original aims of DNA barcoding, including democratization of access to genetics tools and to biodiversity knowledge (Janzen et al. 2005; Holloway 2006), are beginning to be realized. Intensive participation and leadership, both in publications (Figs. 2-4) and in conferences (as reported in
Adamowicz (2015), from several highly biodiverse nations is an especially promising trend from both scientific and societal perspectives.

While substantial progress has been made, challenges to international participation remain. Geary and Bubela (2015) highlight that different communities of researchers and stakeholders may have different priorities and concerns and that strategies should be fostered to enable participation and to minimize conflicts. There is also attention needed to policy development, to enable participation in international research programs while also respecting the need for national protection over genetic resources and access and benefits sharing (Borisenko 2015; Neumann et al. 2015).

Concluding remarks

Increasing global participation in DNA barcoding is a promising trend in genetics and biodiversity research. Global-scale genetic datasets will advance scientific knowledge about the evolution and distribution of biodiversity and will enable diverse applications, including the detection of agricultural pests and invasive species, regulation of trade to ensure food authenticity and safety, and conservation. Moreover, equitable access to participation in scientific research is important for increasing opportunities and for maximizing benefits from human intellectual capacity globally.

Important next steps in documenting global biodiversity in a standardized manner may include increasing the training opportunities available to scientists and establishing more
specimen processing or sequencing facilities, especially to scale up barcoding efforts in the most biodiverse regions of the world (Janzen and Hallwachs, 2011). The formation and maturation of national and regional-scale barcoding networks will also be important in order to increase coverage of DNA barcodes and to increase opportunities for participation in research. For example, the Mexican Barcode of Life represents a successful national-focused network (Escalante et al. 2011; Elías-Gutiérrez and León-Regagnon 2013), which has achieved much despite relatively modest funding inputs (Elías-Gutiérrez and León-Regagnon 2015). Similarly, Argentina has developed a national network and also has offered training workshops and fostered international scientific collaborations, with a focus upon partnerships with other South as well as Central American nations (e.g. Lavinia et al. 2015a, 2015c).

In this opinion article, we presented our view that broad participation in genetics research is beneficial and, furthermore, that DNA barcoding is a promising research domain for facilitating such participation. A consideration of several trends in the published literature generally supported the conclusion that international participation in barcoding is increasing and that greater evenness among nations in authorships is observed in comparison with the broader discipline of genomics. However, true democratization of access to biodiversity knowledge will also require engagements beyond the scientific community, including educational opportunities and technology developments to put access to biodiversity knowledge into the hands of all citizens. The hypothesized affordable hand-held barcoder (Janzen et al. 2005; Holloway 2006) would transform humanity’s relationship with biodiversity. Technological developments are proceeding rapidly (e.g. nanopore sequencing as outlined in Boutain and Boutain 2015, or the single closed tube reaction described by Sirianni and Wangh...
2015), and such progress may outpace the resolution of the geo-political and economic issues that currently limit access to participation in scientific research and access to scientific knowledge more broadly.

Finally, based upon observing the evolution of the research programs of several colleagues and research networks, we also think that participating in barcoding projects can be one path to broader participation in other genetics-related disciplines as well, including metabarcoding, phylogenetics, and genomics. Moreover, for non-molecular biologists, participation in barcoding projects via collaboration can also open new opportunities for research and discovery, e.g. in systematics, ecology, evolution, and conservation biology. Therefore, we advance that DNA barcoding can be a successful medium for increasing research and collaboration opportunities among a broad spectrum of members across the global scientific community.

Acknowledgements

We are grateful to two reviewers for providing thoughtful comments that helped us to improve our paper. Thank you also to Melania Cristescu and to Genome for this invitation to provide our perspective.
References


Genome. 2015. Scientific abstracts from the 6th International Barcode of Life Conference.


Holloway, M. 2006. Imagine a portable DNA barcode scanner that could transform people’s relationship with nature. Could such a futuristic technology be to biodiversity what the printing press was to literacy? Conservation in Practice. 7(2): 14-21.


PLoS ONE 6(8): e18123.


Figure Legends
Figure 1. Growth rate of publications in DNA barcoding (solid line) compared with genomics (dotted line). The percentage of the total publications from 2003-2014 in each discipline that were published in a given year is plotted against year.

Figure 2. Growth rate in the number of nations publishing in DNA barcoding (black bars and solid trend line) compared with genomics (white bars and dotted trend line). The number of total nations represented among the publishing authors within each discipline is plotted for each year between 2003 and 2014. Trend lines show the moving average, with a period of 2.

Figure 3. Distribution of authorships among countries for papers published from 2003-2014 in: A) genomics and B) DNA barcoding. The legend can be associated with each figure by reading the countries from top to bottom and comparing these with the pie slices, moving clockwise and starting at the top. The 6 most highly represented countries among authors are listed. Countries are placed in the same order in both charts to facilitate comparison.

Figure 4. The percentage of total authorships (2003-14) for each discipline is shown for each continent. Genomics is presented by the white bars, with DNA barcoding authorships shown in solid black bars.