INTRODUCTION

iBOL's primary mission is the construction of a DNA-based identification system for all eukaryotes, and the application of this new tool to better manage, discover and protect global biodiversity. The research and administrative activities required to achieve this mission were organized under 5 Themes and 20 Working Groups (WG) whose roles are outlined below. A separate GE[3]LS team has also been established, and its research programs are discussed in Section Vof this progress report.

Theme 1 DNA Barcode Library: iBOL's key scientific deliverable is the assembly of a DNA barcode reference library for at least 500K species. Ten WGs lead this effort, each tasked with setting and achieving targets for species coverage in selected taxonomic groups or environments. WGs 1.1-1.3 focus respectively on vertebrates, land plants and fungi, while WGs 1.4[2]-1.6 ensure that the DNA barcode library includes species of socio-economic importance, such as parasites and disease vectors, agricultural and forestry pests, and pollinators. Finally, WGs 1.7-1.10 are concerned with assembling the barcode libraries needed for bio-surveillance in freshwater, marine, terrestrial and polar environments.

Theme 2 Methods: The four WGs in Theme 2 are tasked with developing protocols that either improve the efficiency or extend the horizons of barcode analysis in important ways.The ability to barcode entire biological assemblages (WG 2.1), to recover DNA barcodes from museum specimens (WG 2.2), to optimize the efficiency and cost-effectiveness of barcode analysis (WG 2.3) and to probe temporal shifts in species composition by recovering barcodes from permafrost deposits (WG 2.4) are all important to iBOL's mission.

Theme 3 Informatics: Two WGs are charged with developing iBOL's informatics platform (WG 3.1) and ensuring its security and functionality on a global scale (WG3.2).

Theme 4 Applications: Two WGs are exploiting opportunities created by completion of the DNA barcode library to better monitor biodiversity (WG 4.1), and to make DNA barcoding a more accessible technology (Mobile Barcoding[3], WG 4.2).

Theme 5 Administration: Two WGs have responsibility for the oversight of iBOL's Project Management (WG 5.1[4]) and Communications (WG 5.2[5]) activities, both of which are critical to establish the research linkages and funding arrangements necessary to sustain iBOL's mission.

The balance of this section of the report reviews the goals and progress of these 5 Themes and 20 Working Groups.

[2] Title of WG 1.4 is changed from "Human Pathogens and Zoonoses" to "Animal Parasites, Pathogens & Vectors ".
[3] Title of WG 4.2 is changed from "Barcoder" to "Mobile Barcoding" to better reflect its goals.
[4] Title of WG 5.1 is changed from "Administration" to "Project Management" to better reflect its purpose.
[5] Title of WG 5.2 is changed from "Outreach and Collaborations" to "Communications" to better reflect its purpose.

THEME 1: BARCODE LIBRARY

Objective

The 2008 ICI proposal for iBOL described plans to assemble DNA barcodes for animals, fungi, and plants from diverse geographic settings. It placed particular emphasis on gaining barcode coverage for taxonomic groups with high economic, ecological or health impacts. The 10 WGs within Theme 1were tasked with the delivery of a barcode library containing records from at least 500K species by December 2015.

Progress

1.1 Overall Progress

Sequence records for the barcode region were available for 97K species in June 2009, representing work conducted over the preceding six years by the DNA barcoding community and 'heritage' data in GenBank gathered as a consequence of general efforts in the sequence analysis of mitochondrial DNA, often for phylogenetic studies. The activation of iBOL has led to dramatic growth in barcode records over the past 18 months with coverage now available for 153K species[6]. These records derive from three primary sources –the Canadian Centre for DNA Barcoding (CCDB), other sequencing facilities allied to iBOL and sequences submitted to GenBank. A Venn diagram (Figure 1.1) reveals limited overlap in species coverage among these three sources. Some replication is desirable, particularly in the case of GenBank records because most of these sequences lack the connection to a voucher specimen required to qualify them for barcode status. This figure makes clear the key role of the CCDB in the generation of barcode data; it has contributed one or more records for 75% of the species (114K of 153K) represented in the current barcode library.

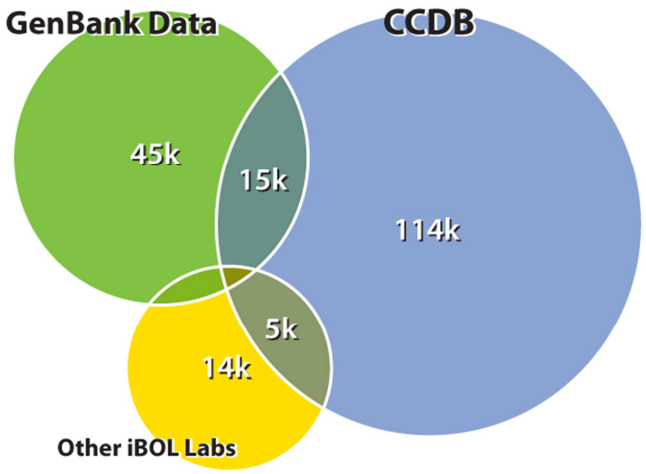


Figure 1.1: Venn diagram showing the three primary sources of barcode data available in December 2010.

From July 2009 - December 2010, the CCDB gathered barcode records from 326K specimens representing 85K different species on behalf of iBOL (Table 1.1). Most (52.5K) of these species represented taxa that previously lacked barcode coverage. Other organizations, many involved in iBOL, contributed first coverage for an additional3Kspecies over this same interval. For example, the MNHM in Paris contributed 8K records representing more than 1K species of marine organisms. As a consequence of these collective efforts, 56K species, 11% of the 500K species target gained their initial coverage in the first 18 months of iBOL operation. Moreover, because of the legacy data, coverage is now present for 153K species (32% of the final target), at a time when one quarter of iBOL's duration has passed. Moreover, the iBOL research plan calls for an increase in activity through time. Reflecting this fact, the budget allocation from Genome Canada for the first 18 months was 16%

[6]Here, we use the term "species" to include both those that have been human-annotated, and those that have been automatically assigned using an *in silico* clustering algorithm.

of its total commitment ($4.3M of $25M) and a substantial proportion of this funding was directed toward a one-time investment in equipment to increase analytical capacity. In summary, overall progress in assembling the DNA barcode library was strong over the first 18 months. As a consequence, the iBOL research team remains confident that the 500K target for species coverage will be achieved by December 2015, particularly since both analytical capacity and specimen acquisition are being scaled up.

Table 1.1a: Number of species in each WG which gained barcode coverage over the first 18 months versus 6 year target.

Working Group		New Species via iBOL	%of Target	Total Species With Barcodes	Species Target (2015)	Progress vs Target
1.1	Vertebrates	1,413	2%	19,018	60,000	32%
1.2	Land Plants	4,663	5%	9,229	100,000	9%
1.3	Fungi	-	-	-	10,000	0%
1.4	Parasites, Pathogens & Vectors	181	2%	1,354	10,000	14%
1.5	Agricultural & Forestry Pests	6,854	28%	17,793	25,000	71%
1.6	Pollinators	5,477	11%	11,162	50,000	22%
1.7	Freshwater Bio-surveillance	2,949	12%	9,448	25,000	38%
1.8	Marine Bio-surveillance	2,424	2%	15,110	100,000	15%
1.9	Terrestrial Bio-surveillance	26,603	27%	64,480	100,000	64%
1.10	Polar Bio-surveillance	1,670	8%	5,009	20,000	28%
	Grand Total	**52,234**	**11%**	**152,603**	**500,000**	**31%**

Table 1.1b: Number of species in each WG which gained barcode coverage over first 18 months versus targets in Notice of Awards from Genome Canada.

Working Group		New species from July 1, 2009 to December 31, 2010	New species targets to June 30, 2011[a]	% Complete
1.1	Vertebrates	1,413	2,500	56.5%
1.2	Land Plants	4,663	6,000	77.7%
1.3	Fungi	-	100	N/A
1.4	Parasites, Pathogens &Vectors	181	320	56.6%
1.5	Agricultural & Forestry Pests	6,854	4,500	152.3%
1.6	Pollinators	5,477	6,000	91.3%
1.7	Freshwater Bio-surveillance	2,949	3,000	98.3%
1.8	Marine Bio-surveillance	2,424	5,000	48.5%
1.9	Terrestrial Bio-surveillance	26,603	26,280	101.2%
1.10	Polar Bio-surveillance	1,670	2,200	75.9%
	Total	**52,234**	**55,900**	**93.4%**

[a] Overall species targets for July 1, 2009 - June 30, 2011 based on GC Notice of Awards.

1.2 Detailed Analysis of Working Group Progress

Tables 1.1a and 1.1b summarize progress for the ten WGs within Theme 1, while Appendix VIII provides a more detailed analysis. The three taxonomically oriented WGs show substantial variation (0- 32%) in progress, largely reflecting differences in the maturity of decisions relating to barcode markers (Table 1.1a). **WG1.1** (Vertebrates) made good progress toward its overall goal with barcodes now available for nearly a third of known species. The need for energized efforts on amphibians and reptiles as well as for work in new geographic regions is being addressed. Barcode records for **WG**

1.2 (Land Plants) are increasing rapidly now that a decision has been reached on the core barcode markers for this kingdom and there is strong confidence that the 100,000 species goal can be reached. The lack of progress for **WG 1.3** (Fungi) reflects the fact that the barcode region(s) for this kingdom have not yet been designated. However, consensus has now been reached for major fungal groups, and papers justifying the designated markers will be published in 2011. Participants in this WG are confident that the original target of barcoding 10,000 species by 2015 remains realistic because enough fungal species are available from culture collections that are active in iBOL to meet this goal.

The three WGs focused on organisms of high socio-economic importance show varied progress, largely reflecting differential success in accessing specimens for analysis. **WG 1.5** (Agricultural and Forestry Pests & Parasitoids) fared best, reflecting specimens generated by general biotic surveys and by pest monitoring programs. **WG 1.4** (Parasites, Pathogens & Vectors) and **WG 1.6** (Pollinators) made less progress toward their targets, but efforts are underway to expand their network of participating researchers. The WGs charged with the development of barcode libraries for ecosystem monitoring collectively made strong progress. **WG 1.9** (Terrestrial Bio-surveillance) reached 64% of its original species goal, prompting this WG to propose a 50% increase in its target to 150,000 species to allow some coverage for groups which were not targeted in the initial proposal. **WG 1.7** (Freshwater Bio-surveillance) and **WG 1.10** (Polar Bio-surveillance) also made strong progress as a consequence of their success in accessing substantial numbers of specimens from diverse taxonomic groups**.** **WG1.8** (Marine Bio-Surveillance) made less progress reflecting:1) the need for further work on primer sets to aid barcode recovery, 2) the inability to use most museum specimens because of their preservation in formalin and 3) the high cost of mounting new collections.

1.3 Overview of Sequencing Activity and Specimen Sourcing

The CCDB extracted DNA from 445K specimens between July 1, 2009 - December 31, 2010 and recovered 326K barcode sequences. The varied iBOL nodes provided 83% of these specimens with Canada making the largest contribution - 18% (Table 1.2). The remaining specimens derived from another 142 countries. The CCDB carried out more than 675K PCR reactions and more than 1.1 million sequencing reactions to generate the 326K barcode records, reflecting the fact that each barcode record is based on a bidirectional read, often of more than one amplicon (e.g. plants & museum specimens). DNA barcodes were not recovered from 117K of the 445K extracts that have passed through the analytical chain (2K extracts remain under analysis). The DNA extracts that failed to deliver barcode-compliant sequences largely derived from aged specimens with degraded DNA. A partial sequence (typically 300-450bp in length) was recovered from 27K of these specimens, but no data were obtained from 90K specimens (20%). Despite the failure to recover sequences from all specimens, barcode records were available for 1.1M specimens by the end of Q6, exceeding the target figure established in the 2009 application by 10%.

The importance of the sequencing activity at the CCDB to the overall iBOL initiative is best indicated by the fact that it gathered 82% of all barcode records generated over the last 18 months.There will be an increased flow of specimens during 2011 and the CCDB is expanding its sequencing capacity to enable their analysis. Researchers with key roles within iBOL are also leading the establishment of major barcode sequencing facilities in other nations. Core facilities with the capacity to analyze at least 100K specimens per year are expected to be functional by 2012 in both China (Kunming Biodiversity Centre, Y-P Zhang) and the United States (National Museum of Natural History, S Miller). Their efforts will be reinforced by facilities that are expected to analyze about 20K specimens per year in India (Aurangabad, JD Khedkar) and Poland (Warsaw, W Bogdanowicz) and likely also Australia and Brazil.

Table 1.2: Number of specimens contributed to iBOL during the period July 2009-December 2010 by researchers in 59 nations. These totals only show the sources of the 326K specimens that generated a barcode-compliant sequence record. The 'Others' category includes specimens from 142 additional nations.

Node	Specimen Count		Other Countries and Regions	Specimen Count
Central	**141,807**		**Other**	**55,296**
Canada	58,864		Indonesia	4,559
United States	48,479		Antarctica	4,121
China	5,956		Gabon	4,064
European Union	28,508		French Guiana	3,167
Austria	*1,251*		Ecuador	3,139
Bulgaria	*1,080*		Thailand	2,036
Finland	*6,983*		Turkey	1,694
France	*2,171*		Pakistan	1,602
Germany	*9,138*		Tanzania	1,463
Italy	*2,532*		Vietnam	1,298
Netherlands	*106*		Bolivia	1,119
Poland	*283*		Taiwan	1,092
Portugal	*247*		Chile	1,026
Spain	*3,303*		Iran	1,022
Sweden	*989*		Comoros	946
United Kingdom	*425*		Belize	923
Regional	**56,732**		Czech Republic	834
Australia	26,783		Paraguay	810
Mexico	12,181		Democratic Republic of the Congo	685
South Africa	5,187		Cameroon	668
Brazil	6,023		Seychelles	665
New Zealand	2,270		Guatemala	593
Norway	1,272		Japan	579
Argentina	1,230		Malaysia	577
Russia	1,360		Philippines	574
India	421		Mozambique	522
Saudi Arabia	5		Others	15,518
National	**71,891**			
Costa Rica	51,151		**GRAND TOTAL**	**325,726**
Papua New Guinea	6,364			
Madagascar	6,083			
Peru	3,348			
Panama	3,029			
Kenya	1,067			
South Korea	614			
Colombia	235			

1.4 Strategic Adjustments Planned For Theme 1

The leaders of the WGs in Theme 1 recognize that success in meeting their targets depends on two factors: 1) the ability of well-established networks of collaborators to provide specimens for sequence analysis and 2) access to centralized sequencing and informatics facilities. Over the past 18 months, these conditions have been satisfied, enabling the generation of a substantial number of barcode records from diverse taxonomic groups, but it has also revealed situations where the supply chain is weak or where there is capacity for increased production. Accordingly, several WG leaders have proposed adjusted species targets and all have identified strategic priorities (Table 1.3). The overall target of gaining coverage from 500K species by 2015 remains unchanged. Strategic priorities that cannot be easily addressed by a single WG, but that require shared effort across multiple groups, will be coordinated through WG 5.1 (Project Management) and WG 5.2 (Communications).

Table 1.3:Planned strategic priorities for Working Groups in Theme 1

Working Groups	**Original Species Target**	**Revised Species Target (proposed)**	**Strategic Priorities & Links to other WGs**
WG 1.1 Vertebrates	60K	10K decrease to 50K	Outreach to barcoding communities for under-represented groups *[WG 5.2 Communications]*
WG 1.2 Land Plants	100K	No change	Access to timely sequencing and informatics resources *[WG 5.1 Project Management]*
			Better primer sets for one of the two plant barcode markers, matK *[WG 2.3 Methods Development]*
WG 1.3 Fungi	10K	No change	Establish barcode loci for different fungal groups *[WG 5.1 Project Management]*
WG 1.4 Pathogens	10K	No change	Funds to support new collection programs, and specimen processing *[WG 5.2 Communications]*
WG 1.5 Pests	25K	5K increase to 30K	Applications of the library for pest management, identifying invasive species, and border protection *[WG 4.1 Environmental Barcoding]*
WG 1.6 Pollinators	50K	15K decrease to 35K	Funds to support collection access and specimen processing *[WG 5.2 Communications]*
WG 1.7 Freshwater	25K	No change	Core facilities to optimize preparation of specimens for barcoding and assurance to collaborating agencies that specimens will be barcoded free. *[WG 5.1 Project Management]*
WG 1.8 Marine	100K	30K decrease to 70K	Improved primers for COI; supplemental markers for certain taxa (e.g. corals, sponges) *[WG 2.3 Methods Development]*
			Funds for specimen collection and tissue sampling *[WG 5.2 Communications]*
WG 1.9 Terrestrial	100K	50K increase to 150K	Expand species target to include *all* groups of terrestrial invertebrates *[WG 5.1 Project Management]*
WG 1.10 Polar	20K	No change	Improve communication between project leaders; recruit new collaborators *[WG 5.2 Communications]*
			Accelerated sequence registration process *[WG 5.1 Project Management]*

THEME 2: METHODS

Objective

The four WGs in Theme 2 are tasked with developing protocols that either improve the efficiency or extend the horizons of barcode analysis in important ways.The ability to barcode entire biotas (WG 2.1), to recover DNA barcodes from museum specimens (WG 2.2), to optimize the efficiency and cost-effectiveness of barcode analysis (WG 2.3) and to generate barcode data from permafrost extracts (WG 2.4) are all important to iBOL's mission.

Progress

WG2.1 Barcoding Biotas

WG2.1 has the primary mission of developing protocols enabling the barcode registration of all multi-cellular eukaryote species from a defined region. Barcoding every species within a particular geographic region presents both significant sampling challenges and a need to develop protocols that support barcode recovery from all taxonomic groups. At the time of project initiation, both locales (Moorea and Churchill) selected as study sites for WG2.1 were thought to host from 5,000 to 10,000 species. Progress at each site has already been sufficient to indicate that the original diversity estimates were too low. For example, more than 5,000 species have barcode coverage at Churchill and there is no evidence of an asymptote in species accumulation over time (Figure 2.1).This result is an exciting one as it reveals the power of DNA barcoding to advance our understanding of biodiversity, but it might also provoke fears that the task of species inventory will never reach completion. In fact, the overall curve conceals the fact that asymptotic values for species diversity have been achieved for a number of intensively surveyed taxonomic groups at both Churchill and Moorea.

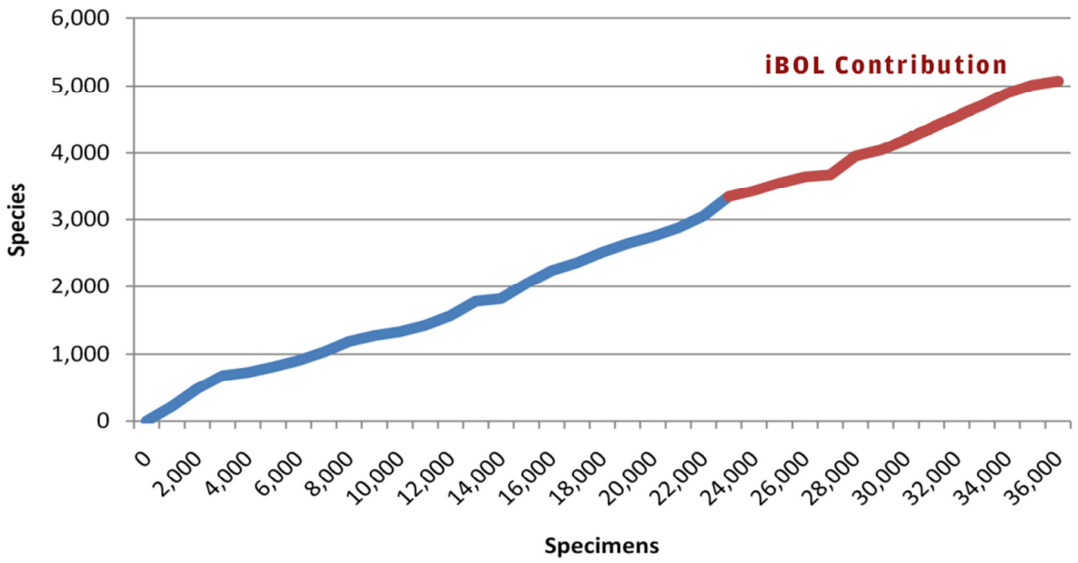


Figure 2.1: The rise in barcode coverage for eukaryotes at Churchill (2007-2010). Progress exceeds that proposed in the 2008 iBOL application which only projected barcode coverage for 5000 species at Q10.

Because of the success at Churchill and Moorea, other iBOL nodes plan to launch similar projects. For example, Argentina and Brazil will barcode the biota of Iguazú National Park, while Mexico is activating projects in the biosphere reserves of Los Tuxtlas and Chamela. Finland is considering a similar effort in old-growth boreal forest, while Norway plans a project in Svalbard, a hot-spot for Arctic research.

WG 2.2 Museum Life

Most (>90%) of the current 1.1M barcode records derive from freshly collected specimens, reflecting past difficulty in barcode recovery from museum specimens because of DNA degradation. WG 2.2 had no quantitative goals in the 2008 iBOL application. However, it was charged with developing protocols that break the barrier to barcode recovery for the several billion specimens held in natural history collections. Past work has established that some of these specimens, especially those stored in unbuffered formalin, present an extreme analytical challenge. However, many museum specimens, including highly diverse groups such as insects and plants, are stored under conditions that favour DNA preservation. WG 2.2 is directing its current efforts towards the analysis of such specimens, investigating how protocols should be adjusted for specimens of varied age. Two large-scale studies are evaluating these matters. The Herbarium at the New York Botanical Garden holds representatives of every genus of land plant and researchers there are testing barcode recovery from these specimens. A second project, involving researchers from CSIRO and the Biodiversity Institute of Ontario, is probing barcode recovery from Lepidoptera held in the Australian National Insect Collection (ANIC). This project has shown rapid progress; 12,500 specimens representing 4,400 species (1/3 of the known Australian fauna) were data based, photographed and tissue sampled in one month by a team of five researchers. Subsequent success in barcode recovery was highly dependent on specimen age (Figure 2.2), but sequences were recovered from more than 95% of the species despite an average specimen age of 24 years. Work is underway to strengthen recovery from old specimens by assembling barcodes from multiple shorter amplicons. However, the early results have been so positive that CSIRO has provided funding from their Transformative Technologies Program to fast track the completion of a barcode library for all Australian Lepidoptera. Their decision has, in turn, motivated curators at the Canadian National Insect Collection to plan similar efforts during 2011 on varied insect orders from Canada. This capacity to better access museum specimens represents an important advance toward iBOL goals for species coverage because it allows a tight restriction on the number of specimens of each species that is analyzed.

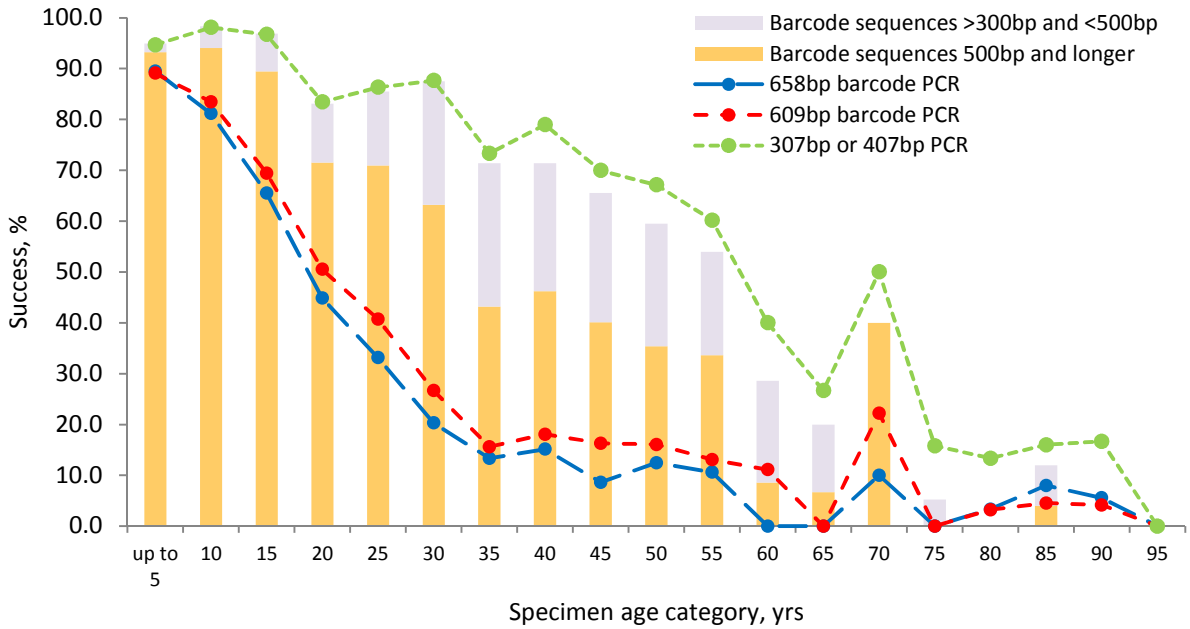


Figure 2.2: Percentage success in PCR amplification for three amplicon lengths (dashed lines; blue = 700bp, red = 612bp, green = 307 - 408bp)for specimens of Lepidoptera of varied age in ANIC. The bars indicate percentage success in sequence recovery for two amplicon lengths with the yellow bar indicating those sequences that qualify as barcodes (length >500bp).

WG 2.3 Methods Development

WG2.3 lacked specific quantitative goals, but it was charged with ensuring that analytical protocols for barcode analysis were honed to remove barriers to barcode recovery and to lower costs. Because of its status as the prototype for future iBOL core facilities, the CCDB serves as the hub for the research activities of WG 2.3. With a staff of 20 research technicians and scientists focused on the generation of DNA barcode data, it provides an ideal setting for methodological innovations that can either reduce the cost or expedite DNA barcode analysis. Following activation of iBOL, the CCDB initiated a substantial revision in operating protocols with the goal of increasing its production from the prior threshold of 100K records per year. These revisions allowed the CCDB to generate 230K records in 2010, but sequencing capacity limited further growth. Two additional ABI 3730XL sequencers were purchased in December 2010, doubling its sequencing capacity. A doubling in the PCR farm (to 50 instruments) and the acquisition of a new high-volume liquid handling system for DNA extraction in February 2011 means that the CCDB now has the equipment to drive production to 400K records a year by 2012. As such, the CCDB is positioned to carry out at least 50% of the sequencing activity required to achieve iBOL's goals. As noted earlier, analytical facilities emerging in other nations will provide the balance of the required capacity.

A Forensic/Troubleshooting unit was established at the CCDB during 2010 to assist other iBOL WGs by developing methods to process specimens that are recalcitrant with standard barcoding protocols. Some of this unit's work has focused on developing novel primer sets for problematic taxonomic groups - fungi (WG 1.3), micro-crustaceans (WG 1.7), rotifers (WG 1.7), and nematodes (WG 1.10). Other studies have been directed toward forensic analyses often involving heavily degraded samples (e.g. food, wood, pelts).

Another major area of activity has been the creation and dissemination of new operating procedures for high-throughput DNA barcoding facilities, including standards for sample submission, and novel methods for the transportation and archival storage of DNA extracts and PCR products. Recent effort has focused on the development of protocols that support barcode-based identifications at a much lower cost to spur application of the technology. By shifting to a unidirectional read and adopting less expensive DNA extraction protocols, it should be possible to drop costs to less than $1. It is critical to emphasize that this analytical chain does not produce barcode compliant records, that it can only be used on freshly collected specimens and that it is not designed to replace existing protocols. It is, instead, designed to support the application of barcode technology in situations where high-volume screening leading to the identification of single individuals is required for regulatory compliance.

WG 2.4 Paleobarcoding

WG 2.4 was charged with developing methods to optimize the recovery of DNA from ancient permafrost cores (>100K years bp). By comparing the sequences of DNA amplified from these extracts with a barcode library from modern taxa, it should be possible to trace shifts in species composition for hundreds of thousands of years. Research in this area has been slowed by the failure to gain support for planned projects in the two nations (Australia, Canada) that were initially charged with leading WG2.4 efforts. Prospects for support within Canada appeared bright as a major application in support of this area of research and other barcode investigations in the Canadian Arctic ($5M over 5 years) advanced through the several stages of review with positive outcomes over a 16-month period, but the application was finally declined in October 2010 because the agency lacked sufficient funding.

Fortunately, iBOL's Regional Node in Norway has gained support and is now leading efforts in paleobarcoding with two major projects. The BarFrost and ECOCHANGE projects are screening DNA from permafrost samples with mini-barcodes to link past shifts in species composition and environmental stability to climate events in order to build models probing the effect of climate change on biodiversity. BarFrost is targeting mosses, fungi, insects, springtails and vertebrates, while ECOCHANGE is focused on vascular plants. These studies benefit in a very substantial way from the barcode library construction carried out by Theme 1, especially its polar WG.

Despite the lack of dedicated funding, the Canadian node of iBOL remains represented in WG 2.4 by the Ancient DNA Centre at McMaster University. Projects underway include the testing of new protocols for the repair of ancient DNA. As DNA ages, damage occurs that impedes the successful application of standard molecular techniques (e.g., PCR). However, new enzymes and molecular protocols aim to repair the DNA, making it available for standard molecular analysis. If successful, this project will not only open the door for the analysis of ancient DNA, but will further enhance the analysis of museum samples (WG 2.2).

Theme 2 Planned Strategic Adjustments

Two of the WGs in Theme 2 are largely based in Canada, reflecting the funding that has been directed toward the establishment of a high-throughput DNA barcoding facility. As new core facilities are established in other nations, researchers at the CCDB will work with their colleagues at these facilities to ensure that methodological improvements are rapidly disseminated. Canada's capacity to sustain its contributions to WG2.1 and WG2.4 are threatened by the failure to gain funds for a major barcode initiative in the Arctic, but other nations are gaining the support needed to progress this work. As mentioned earlier, iBOL strategic priorities cannot be routinely addressed by single WGs, but require shared effort across multiple groups, coordinated through WG 5.1 (Project Management) and WG 5.2 (Communications). Table 2.1 lists strategic priorities for Theme 2.

Table 2.1:Planned Strategic Priorities for Working Groups in Theme 2

Working Groups	Planned Strategic Priorities & Links to other WGs
WG 2.1 Barcoding Biotas	Aid activation of similar projects in other iBOL nodes *[WG 5.2 Communications]*
WG 2.2 Museum Life	Recruit additional participation of museums in iBOL , aiding access to collections *[WG 5.2 Communications]*
WG 2.3 Methods Development	Generate and communicate a "blueprint" for iBOL core facilities worldwide *[WG 5.1 Project Management]*
WG 2.4 Paleobarcoding	Facilitate paleobarcoding projects through new grant applications *[WG 5.1 Project Management]*

THEME 3: INFORMATICS

Objective

The Barcode of Life Data System (BOLD) (www.boldsystems.org) provides storage for DNA barcode records and a data management infrastructure for large-scale barcoding projects. It also includes analytical functions that aid the quality assurance of barcode data and the generation of specimen identifications for projects operating at any scale. The ongoing development and utilization of this resource for the benefit of the iBOL project is the objective of the two WGs in Theme 3.

Progress

WG 3.1 Core Functionality

BOLD has rapidly established itself as the global workbench for the assembly, analysis, and publication of DNA barcode records. The primary goals of this system are to increase the volume, diversity, and quality of barcode data and to provide the tools necessary to maximize the use of these data. Continued growth in functionality, interconnections with other databases, and outreach efforts have been effective in creating rapid growth (Figure 3.1) in both the user base (currently 6K registered users) and volume of data (currently 1.1M barcodes).

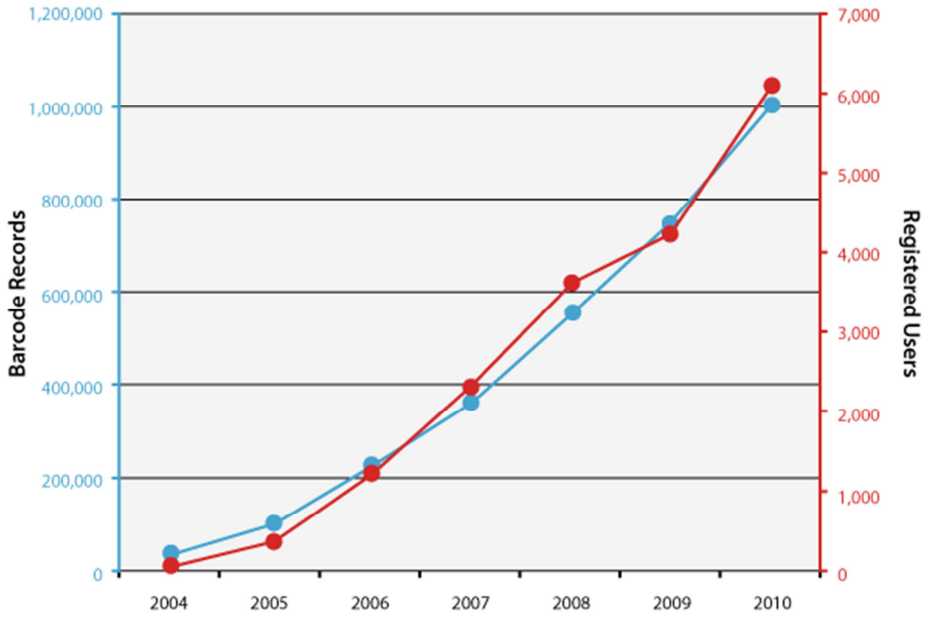


Figure 3.1: Rise in the number of barcode records and registered users in BOLD from 2004-2010.The 2008 iBOL application established the goal that BOLD would house 1M records by the end of Q6, a target exceeded by 10%.

As a result of its active role in the publication of barcode data, BOLD is already a major contributor to the International Nucleotide Sequence Database Collaboration (INSDC), having facilitated the submission of nearly 5% of the eukaryote protein sequences contained in GenBank and it will soon become the majority contributor with respect to taxonomic diversity. Because of its varied functionality, BOLD currently receives 7 million hits per month from 13K unique visitors (Figure 3.2).The rising usage is met through ongoing hardware expansion: 56 processor cores and 16TB of storage were added in 2010, bringing the total to 488 processor cores and 125TB of storage.

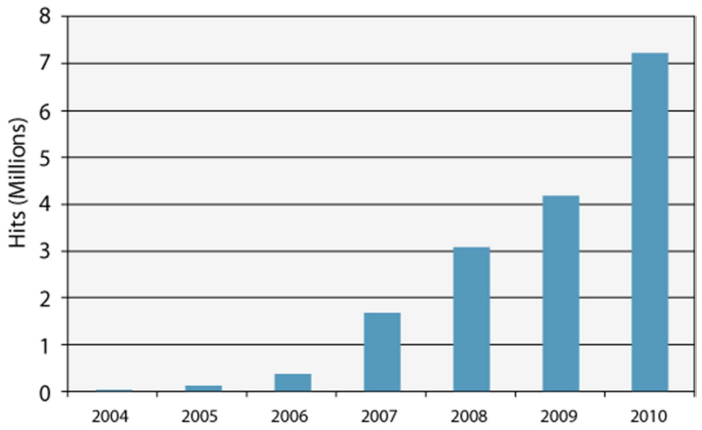


Figure 3.2:Rise in the number of hits on BOLD for the month of October from 2004-2010.

3.1.1 Data Storage, Analytics and Visualization

In preparation for the novel workflows required by iBOL and the high volume of data generated, BOLD has adopted an annual cycle of soliciting input on required functionality from its user community. Requirements from iBOL partners and participants played a large role in the design and rollout of version 2.5 of BOLD, first presented at the 3rd International Barcode of Life conference in November 2009.

The focus of this version was to extend the capacity of BOLD to support DNA barcoding work on plants by enabling the deposition and analysis of the two barcode markers (rbcL and matK) which were selected as plant barcodes earlier that year. However, this reconfiguration of BOLD also provided it with the capacity to support varied ancillary markers (e.g. for fungi). This enhancement has proven useful for investigations that employ supplemental gene regions to examine the status of cryptic taxa revealed through analysis with the primary barcode marker(s).The same functionality has also drawn new users to BOLD, including those involved in multigenic phylogenetic analyses. Figure 3.3 shows the growth in usage of multiple markers over the past year.

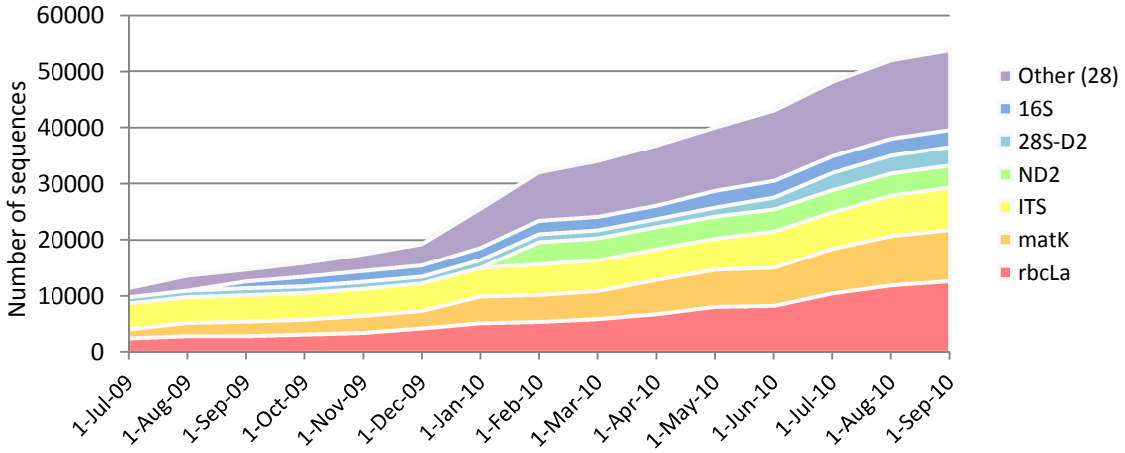


Figure 3.3: Growth of coverage for other gene regions in BOLD.

Aside from its capacity to store sequence records for varied gene regions, BOLD provides analytical tools that help researchers share results and annotate data using a centralized computational resource. This approach has proven very attractive, especially for users that lack the bioinformatics skills or platforms required for large-scale analysis. Over the past year, expansion in this area has focused on developing or extending tools that enhance data quality and that aid the management of a large numbers of records. Four tools are of particular note*:*

Alignment Browser: This interface enables users to review the validity of their alignments, aiding the detection and correction of sequence alignment errors. It provides the functionality of a desktop sequence alignment editor, removing the requirement for third party software that is often difficult for barcode workers in the developing world to afford.

Accumulation Curves: This function generates curves that show the rise in both species representation and barcode clusters with increasing sampling effort. The results of such analysis are highly valuable in helping to adjust sampling effort.

Multi-Marker Analysis: All sequence analysis tools have been updated to consider multiple markers in analysis including the addition of parameters to handle coding and non-coding loci. Some tools have been expanded to aid in the evaluation of marker utility in segregating species.

Real-time QA/QC Reporting: Management consoles have been extended with detailed quality and standards compliance reports generated in real time as data are submitted. Reports are accessible to participants of projects as well as to WG leads who can address emergent quality issues.

3.1.2 Collaboration with GenBank

BOLD and GenBank have been actively collaborating on the development and expansion of a pipeline for the bulk submission and publication of barcode data from BOLD. This pipeline was first used to inject iBOL data in January 2010. This transmission, which involved over 100K records, revealed large gaps in the functionality and process on both sides, largely linked to complexities introduced by differing schema for assignment of GPS values to a particular nation and the use of differing alignment models. Submissions from iBOL continued on a quarterly basis with three batches of 50K records released in April, July, and November 2010. Problems were reduced with each submission and the November submission was nearly error-free (3 issues in 50K records). As the January 2011 submission revealed further cases of alignment discordance, BOLD staff is building a document for submission to GenBank that clarifies the basis of the discords with a view towards resolution. While this matter is under review, BOLD will move to a monthly release cycle for iBOL data. The GenBank flat-files generated from these iBOL submissions contain custom fields displaying the WG and Barcode Index Number assignment for each record (Figure 3.4).This data release process has drastically changed the volume of public barcode data available. User-initiated release through BOLD continues to be active and it has matched the volume of data produced by the broader research community, but the quarterly release of iBOL data has exceeded both of these sources in just one year (Figure 3.5).

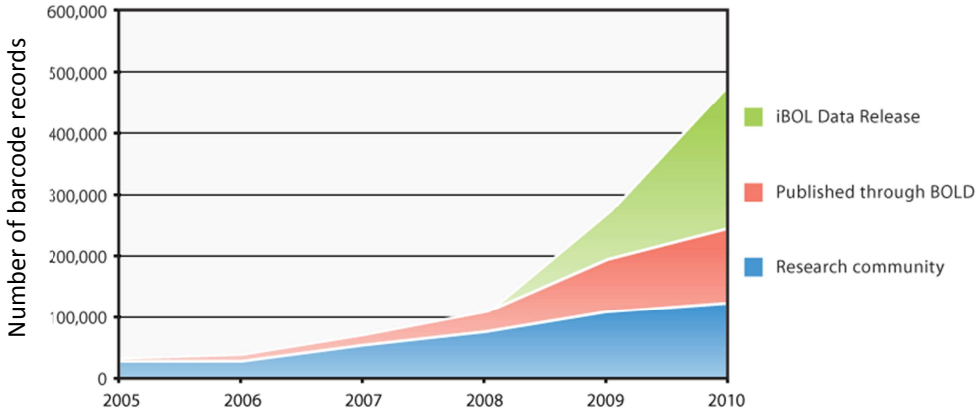


Figure 3.5: Source of barcode sequences in GenBank.

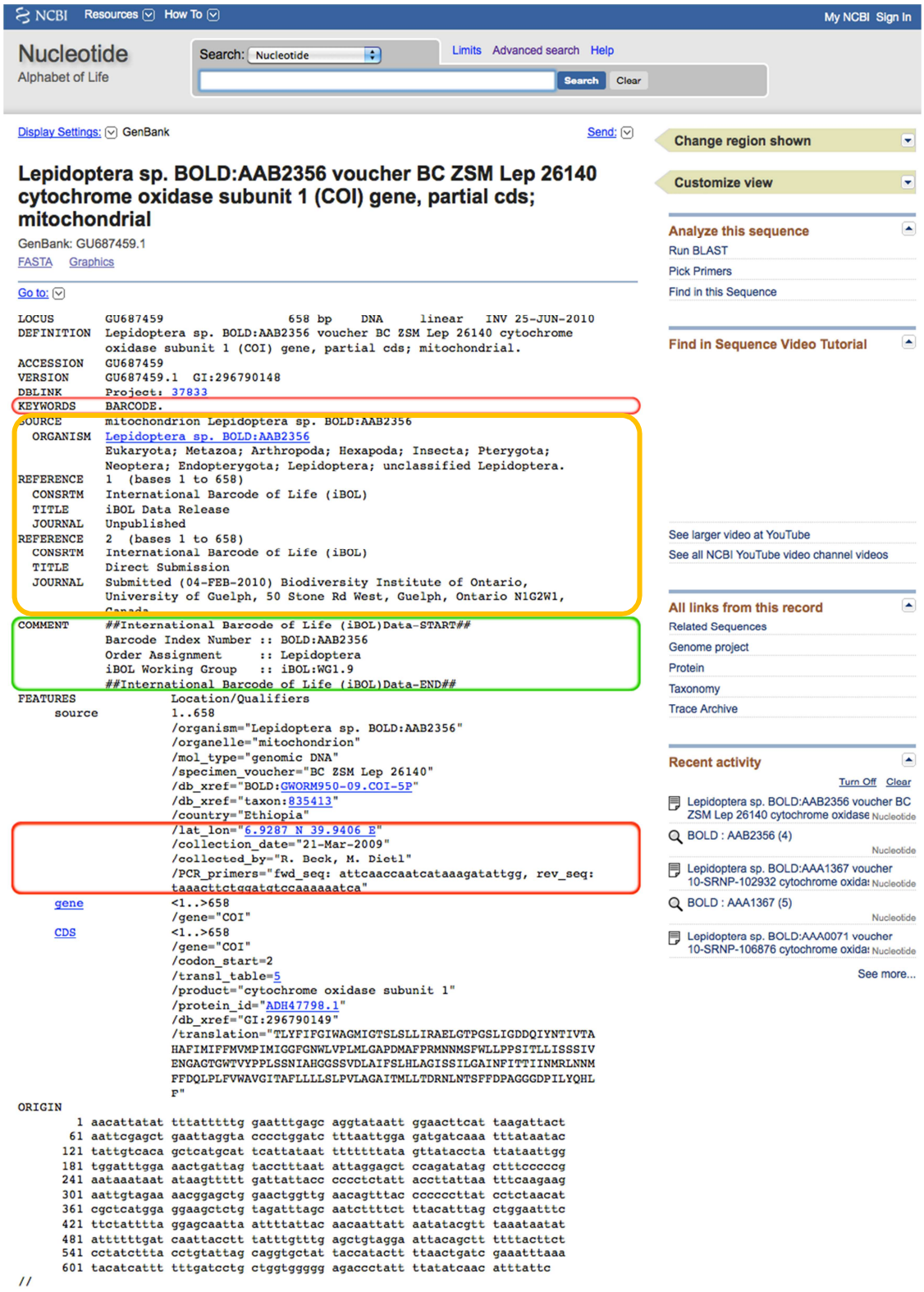


Figure 3.4 GenBank flatfile for a specimen submitted through BOLD. The fields circled in green are those that were added to GenBank as a consequence of iBOL, while those in red reflect fields were added as a result of earlier adjustments made to accommodate DNA barcode standards. The fields circled in orange indicate that the record derived from the iBOL project.

The rapid data release policy adopted by iBOL is a paradigm shift for many users of BOLD. In particular, taxonomic researchers are unaccustomed to releasing data so rapidly (quarterly until January 2011, now monthly). Training efforts on high-speed QA/QC processes, combined with new tools and interfaces, were deployed to address the issue with considerable success. BOLD now has functionality that allows researchers to use tags to indicate barcode records that appear to be flawed as a result of an error somewhere along the analytical chain. Analytical tools have also been improved to provide pre-release reports identifying taxonomic, laboratory, or data entry errors. To ensure the high quality of released data, BOLD staff review and validate all data flowing through the early release pipeline and subsequently resolve any issues that are identified by staff at GenBank.

The release of iBOL data has exposed over 236K COI records on GenBank and on the public side of BOLD since July 2009. To consolidate barcode and ancillary data, a new database has been developed as an extension of BOLD that supports data mining and large-scale review and access to data. Optimized for public data, the new system incorporates micro-attribution for specimen, sequence, and image data, giving proper credit to contributors. Seven requirements are in place for records that are released through this process.

1. Country required and validated against GPS
2. Sequence over 500bp in length with less than 1% ambiguous bases and gaps.
3. BIN assignment present (for animal barcodes)
4. PCR primers present
5. At least two successful trace files present per sequence
6. Order level taxonomy present
7. Sequences must pass reading frame shift tests

Non-compliant records are still publicly released on BOLD and any of these records that gain barcode compliance as a result of further analytical effort (e.g. analysis with additional primer sets) are submitted to GenBank.

3.1.3 Synchronization with GenBank

Steps are being taken to enable the synchronization of any updates that occur on BOLD for records that have been previously transmitted to GenBank. GenBank typically functions as an archival repository for molecular data, maintaining a snapshot of records at the time of publication. Some records undergo subsequent revision by the original submitter due to the detection of errors or new knowledge, but community annotation of sequences is not supported by GenBank. It is important to emphasize that such community efforts are critical to ensure the validity of taxonomic assignments for specimens that served as the source for barcode sequences. Such annotation typically involves iterative inspections of both barcode data and specimens and ultimately comparison with type specimens. To address this need for the barcoding community, BOLD is developing an automated update channel to ensure that all revisions to data on BOLD flow to GenBank. The complexities in the establishment of this channel are considerable and it has required time to marshal the resources necessary for the project. An important early step was the synchronization of controlled vocabularies and data validation protocols. Controlled vocabularies are now posted on a shared FTP site hosted by BOLD starting with synchronized country lists. Audit trails for data submission notifications and queries also follow the same process to ensure proper tracking of issues and transparency of the process. Work on the update channel is in progress and it should be active by May 2011.

3.1.4 Barcode Index Number (BIN) Framework

The BIN system represents a major contribution to biodiversity science that has been developed by the informatics team at BOLD. It employs a clustering algorithm to partition barcode sequences into groups that are treated as putative species. The project was originally intended as a tool to aid rapid data release for organisms that lacked a formal scientific name. However, as implementation progressed, it became clear that the BIN system has the potential to revolutionize the taxonomic process. Researchers given early access to the results of BIN assignments found novel uses for their application in biodiversity surveys. For example, by using BINs as a proxy for species, it was possible to quantify biodiversity patterns in situations where such analysis would otherwise be impossible (e.g. Figure 3.6). When iBOL researchers were first exposed to the BIN concept in November 2009, they welcomed the approach, but urged that formal release of the system be delayed until key functionality was developed to better integrate the system with traditional taxonomic processes. This work

proceeded during 2010, and over this interval, many researchers tested the approach with increasingly positive responses.

The algorithm supporting BIN assignments has now been integrated into BOLD, resulting in the recognition of 153K BINs among its 1.1M barcode records. The first public release of the BIN system, scheduled for February 2011, will include tools that utilize BINs to validate sequence records, to quickly identify cases of taxonomic incongruence, and to compare the results of analyses based on BINs with those from traditional taxonomy. As well, each BIN will be represented online by a page(Figures 3.7 and 3.8 provide examples; a demonstration of the system is available at http://bins.boldsystems.org) which aggregates all knowledge about its members, including:

- All taxonomic names associated with the specimens in each BIN
- All associated images including specimen and habitat photos. The highest quality photograph has been selected for primary display, but all images are accessible
- A distribution map and list of countries with collection sites for all BIN members
- A Neighbour-Joining tree showing all barcode sequences in the BIN
- A list of publications associated with records in the BIN
- A list of attributions for the individuals or organizations responsible for specimen collection, specimen curation, photography, and sequencing

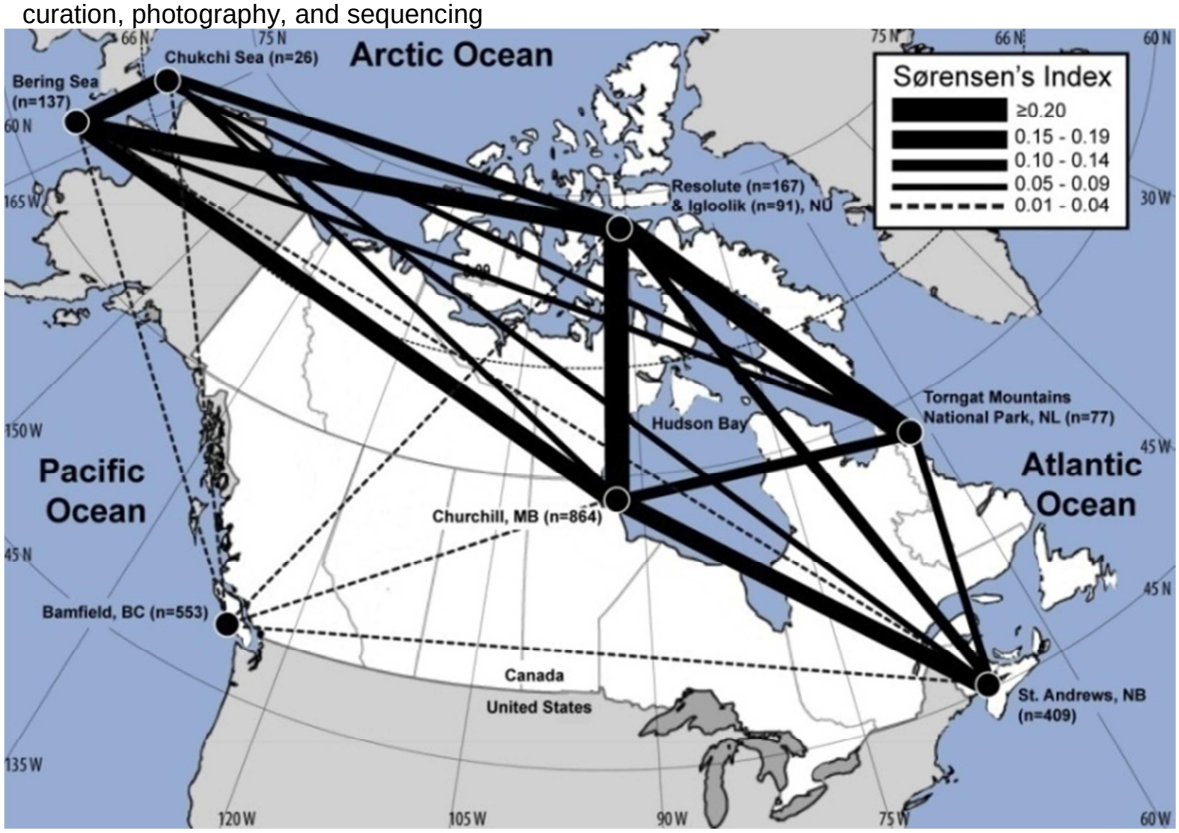


Figure 3.6: Network diagram showing the similarity in polychaete faunas in Canada's three oceans. Sørensen's Similarity Index measures the proportion of species (here represented by BINs) shared between sampling sites. Similarity values (range 0-1 with 1 indicating complete overlap) between sites are represented by line width.

Although the BIN algorithm is very effective for analyzing animal barcode sequences, it cannot easily be extended to plants because the barcode markers for this kingdom show both highly variable rates of nucleotide substitution and alignment difficulties arising from insertion/deletion events. Fortunately, the basic motivation for a BIN system, the presence of many undescribed species, is a much smaller problem for the plant kingdom. Interestingly, protist barcodes are apparently effectively clustered into BINs, but further evaluation by taxonomic experts will occur before the implementation of a BIN system for these organisms.

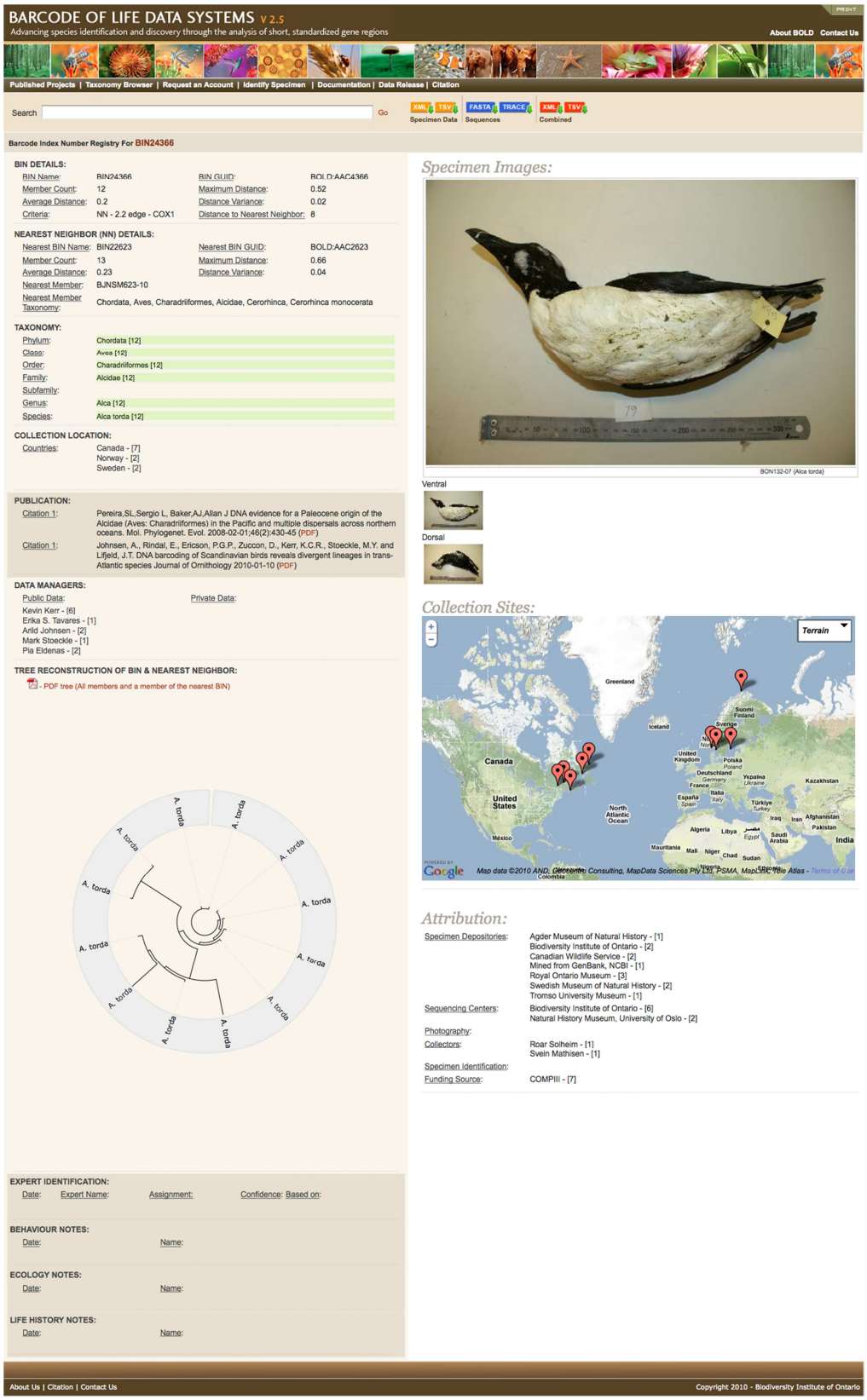


Figure 3.7: Page for BIN:AAC4366, *Alca torda* (Aves: Charadriiformes), a species that occurs along the coastlines of both eastern North America and western Europe. This is one of the 153K BIN pages now available on BOLD.

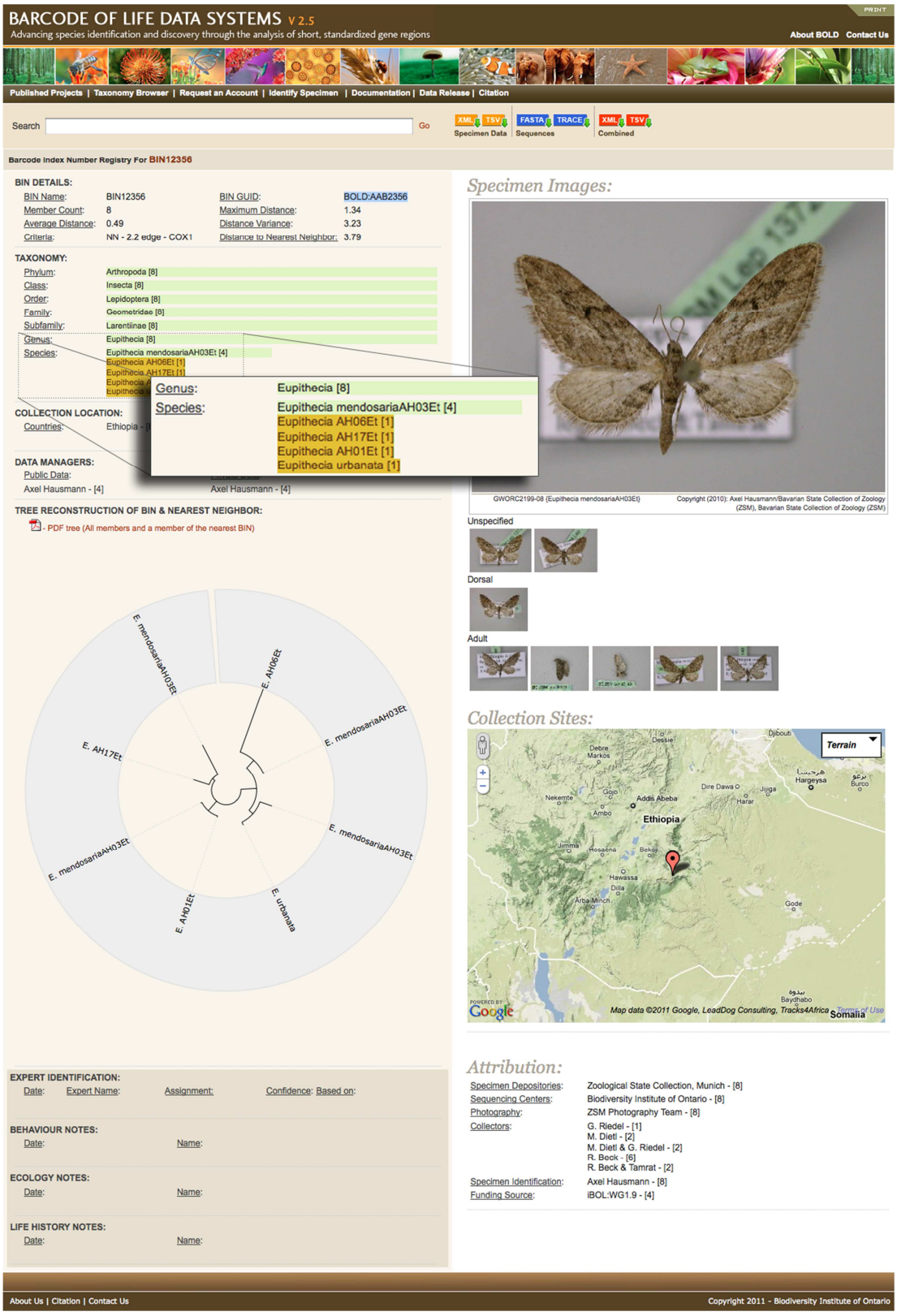


Figure 3.8: Page for BIN:AAA4205, a member of the genus *Eupithecia* (Lepidoptera: Geometridae) endemic to Ethiopia. The identification of this BIN is uncertain; it has been assigned five different names - four interim and one formal, a fact highlighted on the page.

3.1.5Community Engagement and Outreach

Effective training of users and outreach to the scientific community has been important in provoking the adoption of BOLD. Outreach efforts have been extended over the past year, partially to address the increased sophistication of BOLD, but mostly because of the need to engage a growing audience focused on barcode application and integration with other disciplines. BOLD staffers are regularly involved in meetings and conferences, holding the first dedicated BOLD Users workshop in Mexico City in November 2009 which drew more than 150 participants. Aside from training visitors to the CCDB, they have developed the informatics curriculum used in international barcoding workshops and the annually revised BOLD handbook is a key resource for the community.

Cross-platform collaboration and integration requests are increasing and stem from a diversity of projects seeking to link with the BOLD Connectivity Module. From museum collection management databases to taxon-specific community resources, such as FishBase, the rising importance of BOLD on the bioinformatics stage is clear. Although resource constraints have restricted pursuit of some opportunities, BOLD has moved to ensure its broad integration in biodiversity informatics. Four collaborations have led to particularly successful projects:

Genomics Standards Consortium (GSC): Beyond membership in the consortium, BOLD staff has been involved in the development and publication of a new community standard for Minimum Information for an Environmental Sequence (MIENS) that considers barcode data as part of a broader environmental sequence domain.

Encyclopaedia of Life (EOL): BOLD is the primary provider of barcode information to the EOL project, currently delivering information to 153K species pages. Linkouts from EOL as a result of this collaboration account for 2.5% of the traffic to BOLD. Plans are in place to develop bilateral communication modules that would allow researchers to utilize semantic data like body size or metabolic rate from EOL in analyses on BOLD.

DNA Data Bank of Japan (DDBJ): DDBJ maintains a Japanese repository of barcode data developed in direct collaboration with BOLD. Their repository has an identification engine based on BLAST and utilizes the BOLD data structure, the latter having catalyzed the incorporation of a Trace Achive at DDBJ as well.

Global Biodiversity Information Facility (GBIF): Collaboration with GBIF originally focused on their standards for specimen database formats and the use of their records to evaluate the geographic scope of barcode coverage for a particular taxon. Recent activity has led to their hosting of an institutional acronym database (collections registry) that provides the controlled vocabulary system that is employed by BOLD for its organizational registry. Plans for publishing species occurrence data from BOLD to GBIF are also under development.

WG 3.2 Mirrors

The 2008 iBOL application included plans to establish two BOLD mirror sites by Q6 to ensure data security, to share the burden of data entry and to foster the development of novel interfaces and data usage. Organizations in two of iBOL's Central Nodes (China, EU) indicated their early desire to host a mirror site and they have now been joined by one Regional Node (Australia). The Canadian node in iBOL has taken a major role in developing the infrastructure needed for communication and data exchange with the mirror sites, primarily through the establishment of Application Programming Interfaces (APIs). However, the lead organizations in partner nations are playing an increasingly important function in developing tools to support mirror site establishment. The Informatics Division within the Chinese Academy of Science's Institute of Microbiology in Beijing initially focused its activity on the construction of a mirror site for the barcode community in that nation. However, their work has now extended into the development of a software package that will aid the establishment of mirror sites in other nations. This software will see its initial use in 2011 – first by the programming team at the Atlas of Living Australia who will host the mirror site in that nation and subsequently by the EU mirror team in Utrecht. Over the past year, it has become clear that many iBOL nations would like to establish a BOLD mirror that highlights their national contribution to the overall iBOL program and efforts will be made to ensure that this desire is met through efforts overseen by WG5.2.

TABLE 3.1:Functionality upgrades to BOLD from July 2010 - June 2011, largely focused on providing required services to mirror sites. The mirror site in Beijing was activated in August 2010.

Functionality Releases and Updates	Details
JULY 2010	
General Web Service API v1	A web service API was deployed that supports queries and downloads of public specimen and sequence data from BOLD including data mined from GenBank.
HTTP based data files for Mirrors	Data packets were constructed that provide a full snapshot of the public data on BOLD as XML documents for download and importation into any data schema.
DECEMBER 2010	
Trace API v1	The web service API was extended to add a service for downloading trace files in batches. This service supports both block downloads of trace files and on-request downloads. Mirror sites can use the service to download the entire trace database or to retrieve updates on a periodic basis. This first-of-a-kind service raises the profile for provenance data in molecular studies; currently exposing 800,000 trace files (over 400 GB of data).
FEBRUARY 2011	
Image API v1	The web service API will be extended to support downloads of images at multiple resolutions. This service will allow mirror sites to either store retrieved images locally or to retrieve them on demand.
Metadata Profile 2.6 Implementation	BOLD will revise its metadata profile, changes which will impact mirror sites.A new data XML schema will be deployed to allow mirror sites to update their database schemas.
BIN System	The BIN system will see full activation including 153K BIN pages and the associated BIN browser.
Identification Service API v1	An identification service will be provided allowing mirrors to develop their own front end to the BOLD identification engine. This engine will run on the BOLD cluster with 400 dedicated CPU cores, allowing 3,000 identifications per minute. Beyond mirrors, this service will see heavily use due to the growing demand for identifications (currently 20,000 requests/month through copy-and-paste web interfaces).
MAY 2011	
Data Partners Annual Meeting 1	BOLD will host an annual meeting for data partners involving representatives from CBS in Utrecht, CAS in Beijing, and ALA in Canberra. The meeting will establish terms of reference, data sharing policies, and work plans for 2011-2012.

Theme 3 Planned Strategic Adjustments

As the iBOL Project gains momentum, the goals of the WGs in Theme 3 are increasingly being informed by the needs of the other WGs and nodes. Improved project management and communications systems, facilitated by WGs 5.1 and 5.2, will clarify user needs and expectations and allow iBOL to establish MOUs with BOLD, its mirror sites and other providers of informatics support deemed critical to iBOL's mission (e.g. GenBank). Strategic priorities for Theme 3 are listed in Table 3.2.

Table 3.2: Planned Strategic Priorities for Working Groups in Theme 3

Working Groups	Planned Strategic Priorities & Links to Other WGs
WG 3.1 Core Functionality	» Provide iterative feedback to the CCDB and to BOLD to generate user-driven informatics products and services*[WG 5.2 Communication]* » Formalize iBOL MOU with GenBank *[WG 5.1 Project Management]*
WG 3.2 Mirrors	Develop MOUs with mirror sites *[WG 5.1 Project Management]*

THEME 4: APPLICATIONS

Objective

The WGs in Theme 4 are advancing iBOL's interest in two areas of application for DNA barcoding – massive biodiversity scans (WG 4.1) and point-of-contact identifications (WG 4.2) Massive scans of biodiversity will be delivered via new generation sequencing platforms which can deliver large numbers of sequence records from environmental samples. Delivering point-of-contact identifications for single specimens will require a new device that executes all steps in barcode analysis (DNA extraction, PCR amplification, sequencing, reference library matching).

Progress

WG 4.1 Environmental Barcoding

4.1.1 Optimal Use of Next Generation Sequencing Technologies

In collaboration with regulatory agencies such as Environment Canada and the US Environmental Protection Agency, work is underway to develop next-generation sequencing (NGS) technologies for DNA barcode applications in bio-surveillance. Given its accuracy and superior read lengths compared to other NGS platforms, the CCDB has acquired a Roche 454 Genome Sequencer FLX as a platform for environmental barcoding. This work has the primary objective of providing a complete readout of species in bulk environmental samples, such as freshwater invertebrate larvae commonly used to assess the health of aquatic ecosystems through various federal and provincial water quality biomonitoring programs. This approach provides both more accurate and more fine-grained (e.g. species-level) biodiversity information than traditional approaches that relied on morphological sorting by teams of parataxonomists. It can also be extended to other under-studied taxa (e.g. sedimentary meiofauna) and is highly scalable. This work is being advanced by an alliance of genomicists, biodiversity scientists, and bioinformaticians assembled from academia, and various governmental laboratories. It is important to emphasize that the support for this line of research does not derive from the funding provided by Genome Canada for iBOL research, but instead from its Technology Development program.

Although NGS approaches are being harnessed in an increasing number of studies involving prokaryotes, robust and cost-effective use of these systems for environmental assessment programs involving eukaryotes demands a systematic evaluation and optimization of protocols. Using species-rich tropical insects as a model, DNA mixtures pooled from known species in various combinations and concentrations are being queried with NGS workflows. Because standard 454 amplicon workflows can lead to biased results, a two-stage PCR approach has been developed to alleviate this problem. This new methodology can identify nearly all species in a complex mixture with the added potential to quantitatively assess species biomass in a given sample.

4.1.2 NGS and Environmental Barcoding

To demonstrate the potential of NGS analytical approaches, DNA-friendly sampling protocols were injected into the Canadian Aquatic Biomonitoring Network (CABIN) sampling strategy. This shift resulted from a quantitative analysis of DNA degradation due to formalin exposure stemming from prior collecting events. Sampling directly from the new preservation fluid resulted in a simple, non-invasive, cost-effective analytical approach that replaced conventional DNA extraction protocols. These results pave the way for large-scale analysis and will promote rapid uptake of NGS technologies in environmental monitoring. Figure 4.1 provides an example of the information on shifts in species composition that are being gathered using 454 analysis of bulk samples of benthic invertebrates.

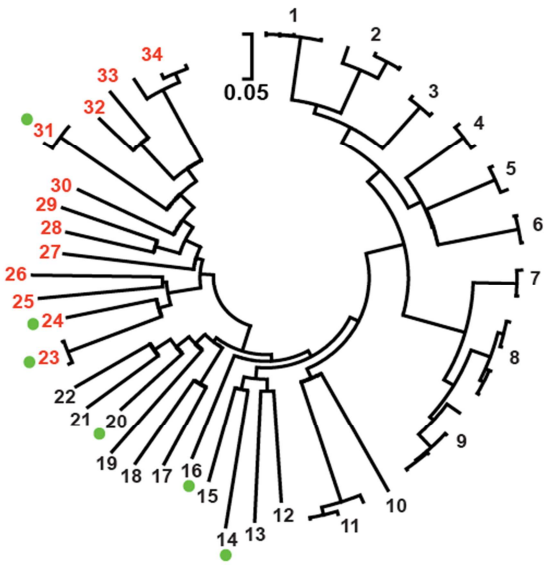


Figure 4.1: 454-Pyrosequencing of bulk benthic larvae can differentiate species of bio-indicator taxa. A neighbour-joining tree of Trichoptera (black numbers) and Ephemeroptera (red numbers) species found in an urban area and an adjacent conservation area. Six environmentally-sensitive species absent from the urban site are indicated by green circles.

WG 4.2 Mobile Barcoding

This WG has not made significant progress, due to recognition during review of the initial iBOL proposal that the development of point-of-contact DNA barcoding is best left to the private sector. However, the well-parameterized barcode library built by iBOL is critical to ensure sufficient application breadth for a return on private sector investment in device development.

Although technical specifications for barcoding on a portable DNA sequencing platform have been discussed with Life Technologies (formerly Applied Biosystems), private sector organizations view such devices as having applications well beyond the DNA barcode community (e.g., in medical diagnostics). Nevertheless, iBOL remains an interested stakeholder in the development of this technology and WG 4.2 will continue to advocate for the development of portable, "field-friendly", and user-friendly DNA sequencing platforms. A key role for iBOL lies in establishing that there is a market for any emergent technology that supports DNA barcode analysis by engaging potential end-users, such as those in federal food inspection agencies and those involved with border inspections.

Theme 4 Planned Strategic Adjustments

Much of the justification for iBOL's work lies in the project's ability to demonstrate successful application of the DNA barcode library for the benefit of biodiversity science and society. There is a compelling strategic need to raise awareness of this theme among all iBOL nodes to stimulate participation in these two WGs. The strong methods development mandate of WG 4.2 also needs strategic linkage to other WGs, particularly those in Theme 2. Proposed strategic priorities for Theme 4 are listed in Table 4.1.

Table 4.1: Planned Strategic Priorities for Working Groups in Theme 4

Working Groups	Planned Strategic Priorities & Links to Other WGs
WG 4.1 Environmental Barcoding	Integrate Methods Development aspects with WGs in Theme 2 *[WG 5.1 Project Management]*
WG 4.2 Mobile Barcoding	Advance awareness, funding and participation *[WG 5.2 Communications]*

THEME 5: ADMINISTRATION

Objective

The purpose of the two working groups in Theme 5 is to orchestrate **Project Management** and **Communications** activities within iBOL's matrix of working groups and nodes. (Figure 5.1, below). As iBOL working groups are globally distributed, with membership and funding drawn from all participating nodes, WGs 5.1 and 5.2 play an important role in linking this matrix to the more centralized resources of iBOL Governance and Management (reported in Section VII) and iBOL core facilities.

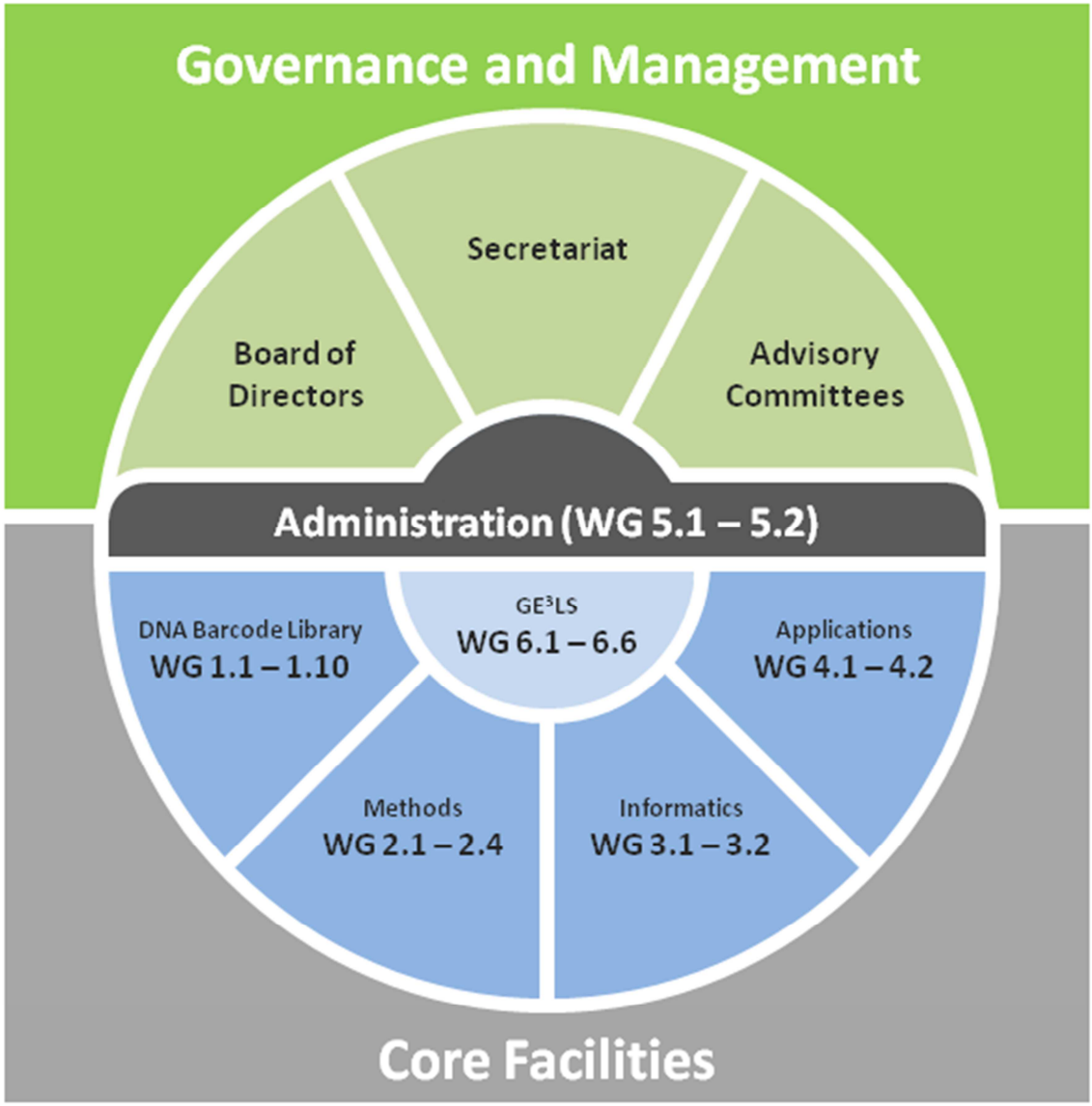


Figure 5.1 iBOL Functional Relationships.

PROGRESS

5.1 iBOL-Canada

In the period under review, the Canadian node of iBOL was the principal actor in (i) mobilizing iBOL's global matrix of Working Groups and Nodes, and (ii) developing and supporting the project's largest sequencing facility, its central informatics platform and its highest-volume specimen processing centre. Progress in these two areas of activity is reported in sections 5.2 and 5.3 below. The Guelph-based administrative unit within iBOL-Canada (Table 5.1) also took responsibility for the financial monitoring and reporting to funders, for activating the Project's Governance and Management structure (reported in Section VII), for collating research contributions (Section VI) and coordinating training and recruitment (Section VIII). Researchers and senior academics from within the Canadian node (including 5 University of Guelph faculty members with a 100% allocation of effort to iBOL) provided leadership to iBOL's public outreach and communication activities, reported in Section IX.

Table 5.1: Members of iBOL Canada's administrative team.

Name	Role
Greg Singer	Project Manager
Susan Mannhardt	Senior Administrative Assistant
Sue-Ann Connolly	Information Officer
Suzanne Bateson	Web /Media Specialist

5.2 Administration of Nodes and Working Groups

5.2.1 iBOL Nodes

In establishing nodes of the iBOL project in other countries, our initial model was for the Canadian node to directly negotiate memoranda of Understanding (MOU) or Letters of Agreement (LOA) with interested parties – both scientific institutions and government agencies – who appeared most likely to be able to build DNA barcoding capacity and advance iBOL's mission in their country or region. Designation of countries as National, Regional or Central nodes (see the sidebar) was explicitly linked to their commitment to provide specimens for barcoding at the iBOL-Canada core facility in Guelph. This model was successful in establishing the 'supply chain' to the iBOL-Canada core facility in Guelph, as evidenced in contribution of specimens to the barcode library, reported elsewhere in the report. There are currently 28 countries listed as iBOL nodes, each with representation on the project's Scientific Steering Committee.

iBOL Nodes

iBOL nodes are the networks of leading researchers and key organizations which oversee capacity building, direction, deliverables and financial accountability for their nation's or region's participation in iBOL. iBOL Nodes whose activities are confined to a single country are designated as **National Nodes**, while **Regional Nodes** are those with the additional capacity to expand partnerships, establish a funding base and develop infrastructure for DNA barcoding and related research on a regional basis that extends beyond their national boundaries. **Central Nodes** are National or Regional Nodes that can and will barcode samples from diverse sources and geographies, and act as leaders in knowledge and technology transfer across (other) Central, Regional and National nodes.

iBOL Working Groups

The purpose of iBOL's working groups (WGs) is to advance iBOL's scientific agenda, which is organized under six themes:

- Theme 1 DNA Barcode Library (WG 1.1 – 1.10)
- Theme 2 Methods (WG 2.1 – 2.4)
- Theme 3 Informatics (WG 3.1- 3.2)
- Theme 4 Applications (WG 4.1 – 4.2)
- Theme 5 Administration (WG 5.1-5.2)
- Theme 6 GE3LS (WG 6.1-6.6)

5.2.2 iBOL Working Groups

Participants in each working group are collaborators from around the world who are engaged in the iBOL Project – either informally as independent contributors, or more formally as part of the Project's evolving nodes structure. Just as collaborators can be mapped to particular geographic nodes, their scientific efforts map them to one or more of iBOL's working groups. As would be expected at this stage of the project, the most heavily populated working groups represent the network of scientists engaged in building the barcode library, whose *de facto* WG efforts can be tracked trough their submission of specimens to the Canadian Centre for DNA Barcoding (CCDB). Membership of working groups in the other themes is more difficult to map, as these working groups will self-select based on research interests and on capacity provided through the nodes structure. All working groups are asked to maintain WG

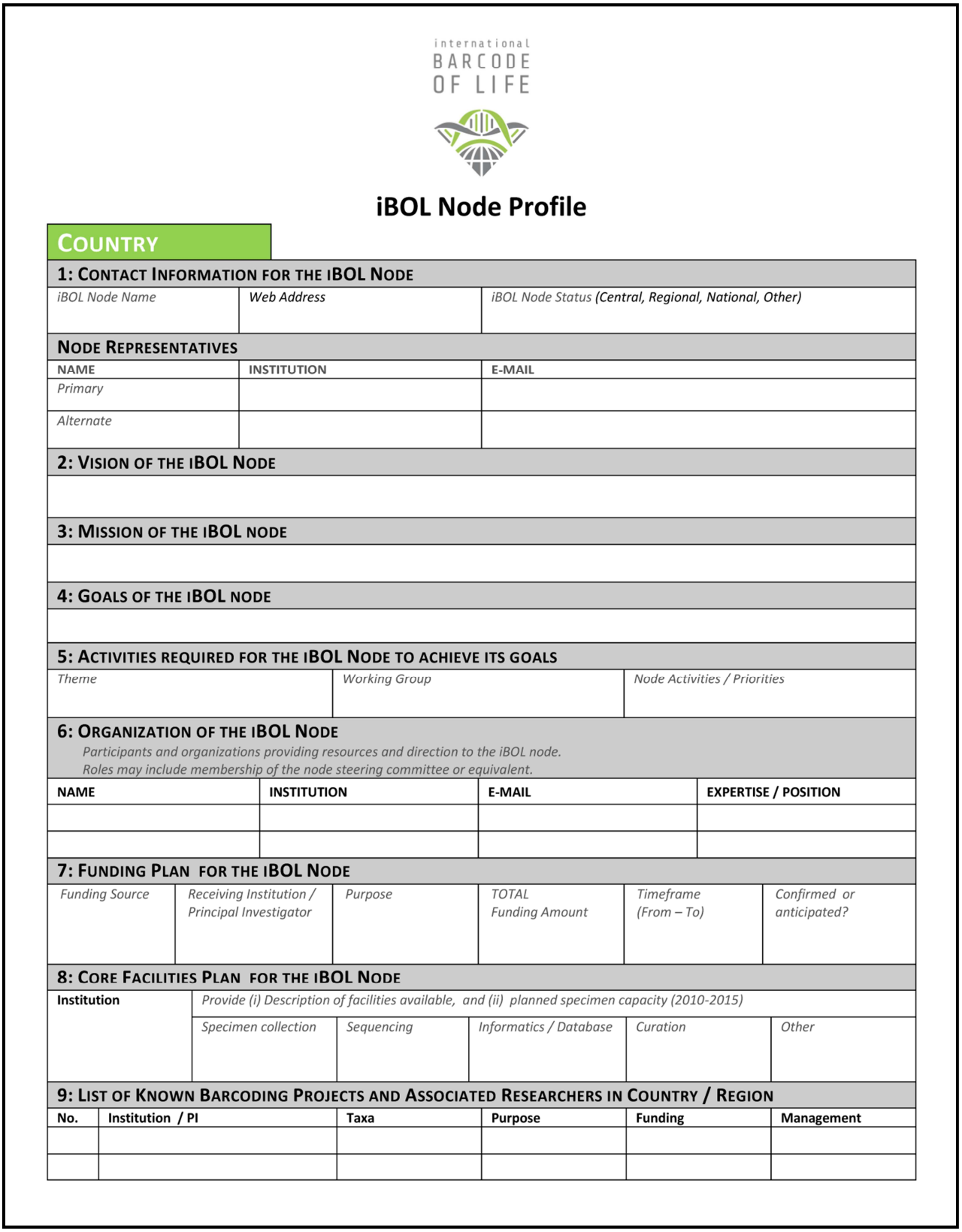
profiles which include information on WG goals, plans, current status and node participation. Profiles for WGs 1.1 – 1.10 are provided in Appendix VIII.

5.3 Administration of Core Facilities

The Canadian node for iBOL supports the project's highest-volume specimen processing centre, its largest sequencing facility(CCDB) and its central informatics platform (BOLD), all hosted by the Biodiversity Institute of Ontario (BIO).As such, the iBOL-Canada node provides an organizational model that is particularly relevant to other Central Nodes. The scale of operations is substantial, as staffing has averaged 56 FTEs since July 2010. These individuals are based in four structural units (Administration - 4, Collections - 13, Informatics - 18, Sequencing - 22) which are collectively responsible for ensuring that iBOL-Canada meets its commitments to the overall iBOL program. The four structural units are each led by a BIO staff member with responsibility for that area of endeavour. The Directors of these units meet weekly, together with the five BIO faculty members and with iBOL's Executive Director and Director of Media and Communications, to discuss both strategic and operational issues.

Planned Strategic Adjustments

The most important adjustment for the two working groups in Theme 5 will be to develop and expand their membership beyond Canada and the US in order provide the necessary global support and coordination of all WGs and Nodes, act as a strategic conduit to and from iBOL's governance and management structure and oversee the development of existing and new core facilities (Fig 5.1). Future MOUs will specify contribution to the work of the administrative working groups, and maintenance of Node Profile information (Figure 5.3) as an expectation of each node's participation in iBOL. Nodes and WGs have also clearly articulated support needs which will be addressed in the plans of WG 5.1and WG 5.2 as the project is advanced (see Tables 5.2 and 5.3 below).



international BARCODE OF LIFE

iBOL Node Profile

COUNTRY

1: CONTACT INFORMATION FOR THE iBOL NODE

iBOL Node Name	*Web Address*	*iBOL Node Status (Central, Regional, National, Other)*

NODE REPRESENTATIVES

NAME	INSTITUTION	E-MAIL
Primary		
Alternate		

2: VISION OF THE iBOL NODE

3: MISSION OF THE iBOL NODE

4: GOALS OF THE iBOL NODE

5: ACTIVITIES REQUIRED FOR THE iBOL NODE TO ACHIEVE ITS GOALS

Theme	*Working Group*	*Node Activities / Priorities*

6: ORGANIZATION OF THE iBOL NODE

Participants and organizations providing resources and direction to the iBOL node.
Roles may include membership of the node steering committee or equivalent.

NAME	INSTITUTION	E-MAIL	EXPERTISE / POSITION

7: FUNDING PLAN FOR THE iBOL NODE

Funding Source	*Receiving Institution / Principal Investigator*	*Purpose*	*TOTAL Funding Amount*	*Timeframe (From – To)*	*Confirmed or anticipated?*

8: CORE FACILITIES PLAN FOR THE iBOL NODE

Institution	*Provide (i) Description of facilities available, and (ii) planned specimen capacity (2010-2015)*				
	Specimen collection	*Sequencing*	*Informatics / Database*	*Curation*	*Other*

9: LIST OF KNOWN BARCODING PROJECTS AND ASSOCIATED RESEARCHERS IN COUNTRY / REGION

No.	Institution / PI	Taxa	Purpose	Funding	Management

Figure VII-3: iBOL Node Profile Template

Table 5.2: Strategic Priorities for WG 5.1 Project Management

Requested by	Strategic Priorities
WG 1.2 Land Plants	Core facilities – access to timely sequencing and informatics resources
WG 1.3 Fungi	Establish barcode loci for different fungal groups
WG 1.7 Freshwater Bio-surveillance	Core facilities – access to allow (i) Optimum preparation of specimens destined for barcoding for the IBOL project; (ii) Assurance to collaborating agencies that specimens supplied for analysis will be barcoded at no charge.
WG 1.9 Terrestrial Bio-surveillance	Expand and refine species targets to include representative taxa for all terrestrial invertebrates
WG 1.10 Polar Life	Core facilities –accelerate sequence registration process
WG 2.3 Methods Development	Generate and communicate "blueprint" for iBOL core facilities worldwide
WG 2.4 Paleobarcoding	Globally facilitate existing and new paleobarcoding projects
WG 3.1 Core Functions	Sign MOU between BOLD and GenBank
WG3.2 Mirror Sites	Sign MOUs between BOLD -Canada and its mirror sites
WG 4.1 Environmental Barcoding	Integrate Methods Development aspects with WGs in Theme 2

Table 5.3: Strategic Priorities for WG 5.2 Communications

Requested by	Strategic Priorities
WG 1.1 Vertebrates	Outreach to researchers to gain access to rare taxa
WG 1.4 Animal Pathogens, Parasites & Vectors	Funds for new collection programs and specimen processing
WG 1.6 Pollinators	Funds for access to collections and specimen processing
WG 1.8 Marine Bio-surveillance	Funds for new collection programs and specimen processing
WG 1.10 Polar Life	Improved communication between projects and project leaders; recruit new collaborators
WG 2.1 Barcoding Biotas	Engage with similar projects in other iBOL Nodes
WG 2.2 Museum Life	Encourage museums to join an iBOL node to aid access to their collections, especially type specimens
WG 3.1 Core Functionality	Strengthen feedback to BOLD to foster user-driven advances in informatics services
WG 4.2 Mobile Barcoding	Advance awareness of the Mobile Barcoding initiative and try to identify funding to progress this effort through dialogue with both governmental and private sector sources.