A Tri-Oceanic Perspective: DNA Barcoding Reveals Geographic Structure and Cryptic Diversity in Canadian Polychaetes

Christina M. Carr1*, Sarah M. Hardy2, Tanya M. Brown3, Tara A. Macdonald3, Paul D. N. Hebert1

1 Biodiversity Institute of Ontario, University of Guelph, Guelph, Ontario, Canada, 2 University of Alaska, School of Fisheries and Ocean Sciences, Fairbanks, Alaska, United States of America, 3 Institute of Ocean Sciences, Fisheries and Oceans Canada, Sidney, British Columbia, Canada

Abstract

Background: Although polychaetes are one of the dominant taxa in marine communities, their distributions and taxonomic diversity are poorly understood. Recent studies have shown that many species thought to have broad distributions are actually a complex of allied species. In Canada, 12% of polychaete species are thought to occur in Atlantic, Arctic, and Pacific Oceans, but the extent of gene flow among their populations has not been tested.

Methodology/Principal Findings: Sequence variation in a segment of the mitochondrial cytochrome c oxidase I (COI) gene was employed to compare morphological versus molecular diversity estimates, to examine gene flow among populations of widespread species, and to explore connectivity patterns among Canada’s three oceans. Analysis of 1876 specimens, representing 333 provisional species, revealed 40 times more sequence divergence between than within species (16.5% versus 0.38%). Genetic data suggest that one quarter of previously recognized species actually include two or more divergent lineages, indicating that richness in this region is currently underestimated. Few species with a tri-oceanic distribution showed genetic cohesion. Instead, large genetic breaks occur between Pacific and Atlantic-Arctic lineages, suggesting their long-term separation. High connectivity among Arctic and Atlantic regions and low connectivity with the Pacific further supports the conclusion that Canadian polychaetes are partitioned into two distinct faunas.

Conclusions/Significance: Results of this study confirm that COI sequences are an effective tool for species identification in polychaetes, and suggest that DNA barcoding will aid the recognition of species overlooked by the current taxonomic system. The consistent geographic structuring within presumed widespread species suggests that historical range fragmentation during the Pleistocene ultimately increased Canadian polychaete diversity and that the coastal British Columbia fauna played a minor role in Arctic recolonization following deglaciation. This study highlights the value of DNA barcoding for providing rapid insights into species distributions and biogeographic patterns in understudied groups.

Introduction

Molecular tools are increasingly recognized as necessary for delineating species boundaries, quantifying diversity, and clarifying distributions in understudied groups [1-3]. These tools can be particularly useful in marine systems where there is a poor understanding of species boundaries and broad-scale distributions. This lack of understanding is driven by the assumption that there are few barriers to gene flow and thus many ubiquitous species [4,5], coupled with the reliance on morphological differences for species recognition [6]. The relatively recent detection of sibling species with restricted ranges, detectable only with molecular tools, supports the use of an integrative taxonomic approach to species delineation and range determination in ocean environments [6,7].

Polychaetes are an abundant and speciose group in marine systems yet they are little studied compared to other taxa of similar ecological importance [8]. Despite their ubiquity (over 10,000 described species [9-11]), polychaetes have generally been excluded from broad-scale distributional studies because of their supposed lack of geographic structure [12-14]. In addition, traditional taxonomic approaches often result in the lumping of phenotypically similar species [15]. Consequently, there are substantial gaps in our knowledge of broad-scale diversity and distribution patterns. However, studies reporting biogeographic structure in polychaetes are becoming more frequent (e.g. along the Chilean coast [16,17] and in the Mediterranean and Black Seas [18]) supporting similar investigation at even broader scales in other regions. Moreover, there is growing evidence that “widespread” polychaete species are more
DNA Barcoding Canadian Polychaetes

often a reflection of the limitations of conventional taxonomy than actual cosmopolitanism, which has led to an increased interest specifically in using genetic data to study taxon distributions [1,19–23].

In Canadian marine waters, nearly 13% of the 1200 known polychaete species are thought to occur in all three oceans [24], but the extent of gene flow among their populations has not been tested. Another 35 species have been recorded in Atlantic and Pacific waters, but not in the Arctic (i.e. have amphiboreal distributions); however, the maintenance of gene flow between modern amphiboreal populations is questionable. In fact, it is possible that species with such disjunct distributions have been isolated ever since their migration through an ice-free Arctic over three million years ago (Ma) [25]. Certainly, extensive genetic structuring has been found in other marine taxa from this region, linked to vicariance caused by the repeated disruption of habitats and dispersal corridors throughout the Pleistocene [26–28]. These factors collectively suggest that a large-scale genetic survey of polychaete species across Canada will reveal overlooked species and shed light on the historic events that shaped modern distributions.

It is now recognized that traditional taxonomic approaches often overlook polychaete species [1] or are confused by colour polymorphisms [29,30]. To address this limitation, several recent taxon-focused studies have examined variation in mitochondrial DNA (mtDNA) sequences and demonstrated that such analysis is valuable for the discrimination of closely related polychaete species [e.g. 15,22,29,31–33]. Studies employing molecular data have consistently split presumed cosmopolitan species into complexes of often very divergent sibling species that have subsequently been found to show variation in other traits (e.g. Eurythoe complanata [15]; Syllis gracilis [19]). The DNA barcode initiative aims to catalogue global biodiversity by linking sequence data to voucher specimens [34,35]. The barcode region for the animal kingdom is a 650 base pair fragment of the mitochondrial cytochrome c oxidase subunit I (COI) gene. Barcoding has frequently been used to recognize provisional species in groups with incomplete taxonomy [e.g. 36,37], and morphological, ecological, and behavioural differences are regularly detected upon further examination of divergent taxa [e.g. 38,39]. Barcoding has been particularly useful in marine systems [5] where variable reproductive forms and life stages, as well as damage to soft-bodied specimens during collection, hinder the identification of even well-known taxa [6]. Moreover, barcodes enable the rapid comparison of multiple taxa from widespread geographic regions and hold enormous potential for assessing genetic cohesion and for detecting broad-scale biogeographic patterns.

While several studies have suggested the efficacy of COI to discriminate polychaete species [e.g. 15,29,33,40,41], no broad genetic investigation, geographic or taxonomic, has been undertaken. In this study, we evaluate the utility of DNA barcoding for discriminating polychaete species, compare richness estimates based on morphological versus molecular taxonomy, examine the genetic structure of widespread species, and investigate broad-scale connectivity patterns among the oceans surrounding Canada.

Results

Amplification success

Three primer sets successfully amplified the barcode region of COI from 1876 of 2324 specimens (81% success). Initial polymerase chain reaction (PCR) with a new primer set polyLCO/polyHCO (Table S1) amplified approximately 70% of the individuals. Most species of the polychaete families Spionidae, Sabellidae, and Cirratulidae were amplified with a primer cocktail [42], while members of the family Serpulidae were amplified with polyLCO/ PolyshortCOIR (Table S1). Species of Serpulidae were amplified with polyLCO/polyHCO using PCR products as template for a second PCR. However, approximately 25 morphospecies, mostly serpulids, failed to yield amplification suggesting residual difficulties in primer binding.

COI genetic distances

Mean COI Kimura two-parameter (K2P) sequence divergence between congeneric clusters was 40 times higher than within-cluster variation (16.50±0.05% and 0.38±0.01%, respectively) (Figure 1). Provisional species (molecular operational taxonomic units, or MOTUs) were identified using a two-step approach (see methods). A first pass screening using a 10× mean intrachuster variation threshold of 3.8% suggested the presence of 315 provisional species. Haplotype network analysis of barcode clusters that were less than 3.8% divergent identified an additional 18 clusters as provisional species. In total, 333 provisional species representing 110 genera and 36 families resulted from analysis of 1876 specimens (Figure S1; Table S2). Presently, 194 MOTUs have been identified to a described species, 97 to a genus level, while the remaining 42 could only be assigned to a family level.

The average number of individuals analyzed per MOTU was 5.7 (range of 1–69), but 119 of 333 were represented by a single specimen. The number of individuals analyzed per MOTU did not affect the average within-cluster variation (Figure 2A); however, maximum within-cluster variation did increase with the number of individuals analyzed per MOTU, though most divergences were under 3% (Figure 2B). Several morphologically identified species had intraspecific divergences of 15–25% (Figure 3A). These taxa always contained two or more distinct genetic clusters, which were partitioned into provisional species with maximum intrachuster variation ranging from 0 to 3.8% (Figure 3B). On average, the maximum divergence within identified species was 5.9%. By contrast, the maximum K2P distance within MOTUs averaged 0.9%. Of the 194 pre-identified MOTUs, 106 had unique species assignments, while the remaining 88 provisional species derived from 34 identified species (Table 1) with an average intraspecific divergence of 12.5%. The commensal scale worms Actinobac fragilis and Actinobac vittata were the only two species to share a barcode cluster. Although intraspecific divergence in this cluster was relatively high (2.81%), the 16 specimens formed a connected parsimony network with no geographic or taxonomic division.

Morphological and molecular richness

Rarefaction curves were generated to compare species richness of morphological and molecular approaches at an equivalent sample size (Figure 4). The barcode curve showed a higher slope than the curve for identified species, with significant divergence (non-overlapping 95% confidence intervals) apparent after 200 specimens were sampled. When the richness of the two curves were compared at the minimum total sample size (n = 1343 morphologically identified specimens), the estimated richness of provisional species was more than twice as high as that for morphologically defined species (295 and 145 species, respectively).

Species distributions

Pacific British Columbia had the highest incidence of provisional species that showed restricted distributions (i.e. were not shared among other sites; 98% of MOTUs), followed by the Arctic (71%), and Atlantic (62%), while the North Pacific (Bering Sea) had the lowest proportion of unique species (48%). Thirty-five percent of provisional species from the Atlantic had distributions...
extending into Arctic waters. By contrast, only 2% of provisional species from British Columbia had ranges that extended into the Arctic. No provisional species showed an amphiboreal distribution. Instead, Pacific and Atlantic populations of amphiboreal species always had more than 13% sequence divergence (Figure 5A; Table 1). Eighteen species with previously reported widespread distributions across Canada were collected from the Pacific and either or both the Atlantic and Arctic Oceans. Fourteen of these species were split into provisional species with large genetic divergences between Pacific and Atlantic-Arctic lineages (Figure 5B; Table 1) and five species had multiple divergent lineages (Figure 5C; Table 1). Eight provisional species had distributions spanning from the Bering Sea to the Eastern Arctic or Atlantic, but only one, *Harmothoe imbricata CMC01*, showed a continuous amphiboreal-arctic range extending from British Columbia to New Brunswick (Table 1).

**Faunal connectivity between oceans**

The highest similarity in species composition occurred between Churchill and the Bering Sea (0.49) followed by Churchill and Nunavut (0.41), while the lowest congruence was between British Columbia and all other sites (mean = 0.05; Figure 6). Connectivity between sites in the Arctic Ocean and North Pacific (Bering Sea) was nearly seven times higher than connectivity between Arctic and boreal Pacific (British Columbia) collection sites (mean = 0.40 versus 0.06, respectively). Comparing the Arctic with boreal faunas of Atlantic (New Brunswick) and Pacific (British Columbia) basins, overlap was two times higher with the Atlantic (mean = 0.10 versus 0.05, respectively) (Figure 6). The lowest within-basin connectivity occurred in the Pacific, between British Columbia and Bering Sea sampling sites (0.09; Figure 6).

**Discussion**

The present study demonstrates the effectiveness of DNA barcoding as a tool for species identification in polychaetes. The clustering pattern of COI barcodes flagged misidentifications, guided taxonomic decisions, and facilitated the detection of diversity overlooked by the current taxonomic system. Of the 142 species morphologically identified in this study, 34 contained multiple lineages representing nearly three times as many provisional species. This result suggests that broad application of DNA barcoding will accelerate the recognition and subsequent description of many currently undescribed polychaete species. Additionally, barcodes enabled the rapid assessment of species distributions, highlighted biogeographic trends, and provided insight into the events that have shaped the contemporary distributions and biodiversity of Canadian polychaetes.

**A DNA barcode reference library for Canadian polychaetes**

This study involved the analysis of nearly 2000 specimens, representing 333 provisional polychaete species. Taking into account the 2:1 ratio of provisional to identified species detected in this study (Figure 4), about 14% of the 1200 known Canadian species [24] have been surveyed. The Arctic was the most heavily sampled region with 140 provisional species (13% of known fauna), followed by the Pacific (120 provisional species, 8% of known fauna), and the Atlantic (86 provisional species, 9% of known fauna). Within the BOLD (Barcode of Life Data Systems) [35] reference library, sequences can be matched to 106 species supported by morphology and barcodes, 34 identified species representing 88 provisional barcode-supported species (e.g. *Pectinaria granulata CMC01*), and 97 genus-level provisional species (e.g. *Nephys sp. CMC01*). Although several provisional species lack formal description, barcode results provide useful information about species’ distribution and ecology as well as a framework for future taxonomic work [43]. Many of the species identified in this study are widespread throughout the Arctic and North Atlantic, suggesting the immediate utility of this newly populated sequence and voucher specimen library for the identification of specimens collected throughout these regions.
Barcode clusters as provisional species

In most cases, the delineation of species boundaries from barcode data was straightforward. Although named species often included more than one provisional species, clusters were easily identifiable and typically highly divergent from other clusters. Barcode sequences differentiated all but one pair of morphologically identified polychaete species. The highest observed within-cluster divergence was 3.8%, indicating that barcodes naturally form tight clusters with low variation. Moreover, the 40-fold higher mean sequence divergence between than within clusters and the rarity of intermediate divergences (Figure 1) indicates that COI barcodes have high discriminatory power for polychaetes. For example, of the 333 recognized provisional species, only 18 showed <4% divergence from another cluster. However, this study also revealed sibling taxa that require more detailed investigation. For example, the maximum divergence within a lineage exceeded the minimum nearest-neighbour distance in two cases: Harmothoe imbricata CMC04 (3.1% and 2.6%, respectively) and Pholoe sp. CMC03 (2.8% and 2.1%). These patterns in variation might reflect the young age of recently isolated populations in North Atlantic and Arctic regions, which are known to harbour species with complex genetic patterning (e.g. Macoma balthica) [28]. Recent origins coupled with complex historical distributions make such young species particularly difficult to diagnose [7,44].

Levels of between-species sequence divergence reported here are consistent with those found in other polychaete studies [e.g. 31–33,45,46]. Prior work has established that cryptic polychaete taxa discovered through genetic analysis regularly show correlated variation in other traits. For example, lineages of the Polydora cornuta complex [33] are reproductively isolated, members of the Hediste japonica complex show differences in life history [47], and sympatric sibling species of the Polydora ciliata complex show ecological differences [48]. Similar indications of ecological divergence were detected between provisional species delineated in this study. For example, habitat specialization was apparent in
several taxa (e.g. Hudson Bay Eteone and Ophelia, British Columbia Syllidae). In other cases, high intraspecific divergences were detected between allopatric lineages (Figure 5; Table 1). Results suggest that populations of species assumed to have distributions spanning Canada’s three oceans have gained reproductive isolation. This is likely a consequence of longstanding range fragmentation caused by cycles of glacial advance and retreat during the Pleistocene, which ultimately promoted diversification in northern marine taxa [e.g. 26–28]. While reproductive isolation is difficult to test among allopatric taxa [7], it can be inferred from molecular divergence between sympatric sibling taxa [49], or from persistent population structure in vicariant lineages following secondary contact [50]. Cases of both phenomena were apparent in this study. Divergent clusters within a species were evident in the absence of geographic barriers (e.g. British Columbia Naineris dendritica). Similarly, six lineages in the Harmothoe imbricata complex have come into secondary contact in the central Canadian Arctic, but remain distinct [51]. Whether these relatively young provisional species will continue on separate evolutionary trajectories or eventually merge as a result of introgressive hybridization requires further analysis of populations in zones of sympatry. Evidence for cryptic species in one geographic region suggests that a ‘phase two’ study linking previously undocumented diversity with ecological and historical correlates might reveal novel insight into mechanisms of speciation and historical phylogeography in polychaetes.

Comparison of morphological and molecular species richness
Barcode-based species accumulation curves for Canadian polychaetes were considerably steeper than those based on morphologically delineated species (Figure 4). After analysis of 300 specimens, genetic species richness was 1.5 times higher than...
morphological species richness, and after analysis of 1000 specimens, richness estimates were doubled. Treating provisional species as functional units of biodiversity, these trends suggest that species richness in Canadian polychaetes is presently underestimated by a factor of at least two.

The effectiveness of DNA barcoding was originally established through work on groups where a strong taxonomic framework enabled tests of the validity of barcodes. The opportunity to critically examine whether COI-delineated species reflect morphological species in polychaetes occurred in the families Nereididae and Polynoidae, which were identified by taxonomic experts. In differentiating species, most disagreements between diagnostic methods reflected a conservative morphology. All identified Nereididae specimens were correctly assigned to species, no species shared a barcode, and divergences between species far exceeded those within. Barcodes further partitioned members of two species into geographically distinct lineages: the widespread Nereis pelagica (average 3.7% divergence), and Alitta virens (15.56%) (Figure 5; Table 1). Further study revealed morphological differences in the provisional Pacific species, Alitta sp. CMC01, which is currently under taxonomic investigation. Barcodes also flagged possible cryptic diversity in three species of Polynoidae: Harmothoe imbricata, Harmothoe rarispina, and Lepidonotus squamatus, all of which exhibited deep genetic divergences between Pacific and Atlantic lineages (Figure S1; Table 1). Additionally, in the Polynoidae, barcodes flagged two genetically indistinguishable species in the genus

| Species                      | No. MOTUs | % K2P | Lineages | Type locality          |
|------------------------------|-----------|-------|----------|------------------------|
| Alitta virens                | 2         | 15.6  | a        | Norway, AT             |
| Barantolla americana         | 2         | 22.9  | a b      | California, PC         |
| Brada villosa                | 2         | 20.1  | a b      | Norway, AT             |
| Capitella capitata           | 2         | 21.8  | a b      | Greenland, AT          |
| Chone magna                  | 2         | 18.9  | a b      | California, PC         |
| Cirratulus cirratus          | 3         | 18.2  | a b c    | Europe, AT             |
| Eteone longa                 | 4         | 13.7  | a b c d  | Spitsbergen, AO        |
| Euclymene zonalis            | 2         | 16.9  | a b      | Maine, AT              |
| Euclidonta minuta            | 2         | 4.1   | a b      | Davis Strait, AO       |
| Eunoe nodosa                 | 2         | 2.6   | a b      | Norway, AT             |
| Eunoe oerstedii              | 2         | 5.5   | a b      | Sweden, AT             |
| Flabelligera affinis         | 2         | 16.5  | a b      | Norway, AT             |
| Harmothoe imbricata          | 6         | 11.6  | a b c d e| North Sea, AT          |
| Harmothoe rarispina          | 3         | 10.2  | a b c    | Norway, AT             |
| Laonice cirrata              | 2         | 19.4  | a b      | Norway, AT             |
| Leitoscloplos pugettensis    | 5         | 15.9  | a b c d e| Washington, PC         |
| Lepidonotus squamatus        | 2         | 14.8  | a b      | Europe, AT             |
| Lumbrineris fragilis         | 2         | 3.0   | a b      | Denmark, AT            |
| Myxicola infundibulum        | 2         | 15.8  | a b      | Medit. Sea, AT         |
| Naineris dendritica          | 4         | 5.8   | a b c d  | British Columbia, PC   |
| Neoamphitrite robusta        | 2         | 11.6  | a b      | Washington, PC         |
| Nephys punctata              | 3         | 8.5   | a b c    | Alaska, PC             |
| Nerephylla castanea          | 2         | 22.7  | a b      | Japan, PC              |
| Nereis pelagica              | 3         | 3.7   | a b      | Europe, AT             |
| Nothria conchylega           | 2         | 6.6   | a b      | Norway, AT             |
| Ophelia limacina             | 2         | 8.7   | a b      | Norway, AT             |
| Ophelia acuminata            | 2         | 3.8   | a b      | Denmark, AT            |
| Pectinaria granulata         | 2         | 16.4  | a b      | Europe, AT             |
| Phyllocoelus groenlandica    | 2         | 13.0  | a b      | Greenland, AT          |
| Praxillata praetermissa      | 2         | 3.5   | a b      | Norway, AT             |
| Scalibregma inflatum         | 2         | 5.5   | a b      | Norway, AT             |
| Syllis alternata             | 5         | 26.7  | a b c d e| Alaska, PC             |
| Syllis elongata              | 2         | 4.2   | a b      | Washington, PC         |
| Terebellides stroemi         | 4         | 20.3  | a b c d  | Norway, AT             |

Each MOTU in an existing species is assigned a letter code (a-f). For cases in which species were split into more than two MOTUs, Kimura two-parameter (K2P) sequence divergence was averaged between all lineages.

doi:10.1371/journal.pone.0022232.t001

Table 1. Species with multiple MOTUs across Atlantic (AT), Arctic (AO), and Pacific (PC) Oceans of North America.
Arctonoe. Barcode sharing may arise in a variety of ways [52], but hybridization is possible since gametes of these species are highly compatible in reciprocal cross studies, suggesting their potential for gene exchange [53]. These species occur sympatrically and morphological intermediates were observed in this and other studies [53,54]. Intermediate phenotypes are a general feature of hybrids [53], but may also indicate that A. vittata and A. fragilis constitute a single species with high phenotypic plasticity. Further examination with various nuclear genes should aid clarification of species boundaries in this genus.

Species distributions and oceanic connectivity

A recent compilation of polychaete species recorded from Canada indicated that 148 species (12% of the fauna) occur in all three oceans surrounding Canada [24]. However, high divergence within several of these widespread species (Table 1) suggests that true cosmopolitanism may be much less common than previously reported. Most species with amphiboreal-arctic distributions examined in this study contained multiple provisional species with enough genetic divergence to designate them as provisional species (Figure 5; Table 1). Specifically, most species that transverse the Bering Strait with ranges extending from the Pacific to the Arctic or Atlantic only occur in the cold waters of the Bering Sea, and do not continue south into boreal British Columbia. Instead, large genetic discontinuities were consistently detected between British Columbia and Arctic or Atlantic populations, suggesting their long-term separation (Figure 5; Table 1). Connectivity patterns suggest that Bering Sea and Atlantic regions served as refugia during Pleistocene glaciations and that these faunas played a key role in Arctic recolonization following the retreat of ice. Conversely, the lack of similarity among coastal British Columbia and Arctic sites suggests that this fauna survived in a more southern Pacific refuge and did not contribute substantially to post-Pleistocene Arctic recolonization (Figure 6).

Divergences among lineages in this study may be attributed to historical range expansion through the Arctic Ocean via the Bering Strait and subsequent isolation when glaciers set in and water levels lowered. The northward flow of Pacific water through the Bering Strait provides the only natural connection between Pacific and Atlantic Oceans in the Northern Hemisphere [56]. Although many taxa likely originated from Pacific ancestors that migrated eastwards into the North Atlantic [27,57], most were originally described from the Atlantic (Table 1) suggesting new names for Pacific lineages may be required. Consistent with other studies [e.g. 15], there was no evidence for contemporary gene flow among Atlantic and Pacific populations of amphiboreal species as no provisional species showed this type of distribution and all morphologically identified amphiboreal sister taxa were genetically distinct (Figure 3A; Table 1). Many temperate species that are thought to occur in boreal Atlantic and Pacific regions are likely distinct, reflecting their separation since a trans-Arctic migration 3.5 Ma [25,58] or since the formation of ice sheets in North America 2.5–3 Ma [59]. The discovery of cases in which genetic data support the splitting of morphological species into multiple provisional species was anticipated, given the high incidence of cryptic species in polychaetes [1,22,60] and the complex glacial history of North America [25–27].

Consistent with other marine taxa [25,58], the highest proportion of unique species was apparent in the Pacific (British Columbia), while Arctic and Atlantic species typically had broad distributions (Figure 6). Interestingly, the Atlantic had a lower proportion of unique species in this study than the Arctic, which may reflect its substantial impoverishment during the last glaciation and its recolonization by Pacific and European species [25,58]. The lowest proportion of unique species was observed in the Bering Sea fauna. The Bering Sea region is a probable refuge for Pleistocene taxa [61], and the high connectivity with Arctic fauna suggests that this region was an important source for today’s Arctic species. Indeed, many species in this study ranged from the Bering Sea through the Arctic and into the Atlantic. Furthermore, the highest overlap in species composition was observed between the Bering Sea and Hudson Bay. The Hudson Bay fauna is an impoverished subset of Arctic species [24], suggesting that many Hudson Bay polychaetes originated from the Bering Sea region. High connectivity between the Bering Sea and Arctic regions reflects recent or newly reestablished gene flow across the Bering Strait since its most recent opening ca.14,000 years ago [62] or periodically throughout the Pleistocene. Such recent gene flow
between Pacific and Atlantic-Arctic waters has also been noted in echinoderms [63], algae [64], and molluscs [28].

This study corroborates the divergent nature of British Columbia and Bering Sea faunas first noted by Ushakov [65], who reported a biogeographic division between Pacific polychaetes on Asiatic and North American coasts. Considerable genetic structure in the Northeast Pacific has also been documented in the gastropod *Nucella lamellosa* [66] and the sea cucumber *Cucumaria pseudocurata* [67], suggesting Pleistocene survival in both northern and southern refugia [66,67]. Ocean currents and temperature are thought to reinforce the break in North Pacific species distributions. For example, anticyclonic circulation of water in the Bering Sea and the porous Aleutian Island boundary may limit southward dispersal of Bering Sea species into the Gulf of Alaska [56]. Additionally, the faunal disconnect in the Pacific has been linked to the influx of Arctic waters into the North Pacific via the Bering Strait beginning 3–4 Ma [59], which promoted the division of boreal faunas on East and West Pacific coasts [57,65]. Subsequent warming of coastal British Columbia may have forced cold water species north into the Bering Sea region [65]. Further sampling of populations between Arctic and boreal regions will aid in determining the approximate location of this well-supported biogeographic boundary in the Pacific.

**Conclusions**

DNA barcoding provides an effective approach for the rapid evaluation of species richness in groups where many species await description. Furthermore, it enables the mapping of species distributions in the absence of formal taxonomic descriptions. The present investigation revealed the likely presence of many undescribed polychaete species in Canadian waters. This discovery was not unexpected since the number of polychaete species that await description is high. This study provides further support that DNA barcoding can accelerate the detection of overlooked species while creating a detailed taxonomic scaffold to aid future taxonomic research. Results of this study also support the assertion that far fewer polychaete species are as widely distributed as previously thought. Expanding the DNA reference library of northern hemisphere marine life holds enormous potential for detailed insights into broad-scale biogeographic patterns. This will ultimately lead to a better understanding of biodiversity, barriers to gene flow, and the process of speciation in ocean environments.

**Figure 5. Examples of geographic splits in widespread species.** Neighbour joining (NJ) trees with K2P divergences between amphipod species with divergent Pacific and Atlantic lineages (A), between species with divergent Pacific and Atlantic-Arctic lineages (B), and between species with multiple divergent lineages (K2P divergence is averaged among the three lineages) (C). Red, black, and yellow shaded regions correspond with branch colour on the NJ tree and indicate different lineages.

doi:10.1371/journal.pone.0022232.g005
Materials and Methods

Specimens

A total of 2324 polychaetes were collected (detailed specimen information is available in the “Polychaetes of North America (PONA)” project on BOLD; www.barcodinglife.org [35]) between 1999 and 2009 from eight locations in Canada and Alaska. These sites included Bamfield (British Columbia; n = 553), Bering Sea (Alaska; n = 137), Chukchi Sea (Alaska; n = 26), Churchill (Manitoba; n = 864), Resolute (Nunavut; n = 167), Igloolik (Nunavut; n = 91), Torngat Mountains National Park (Newfoundland and Labrador; n = 77), and St. Andrews (New Brunswick; n = 409). Samples from most sites were collected in the intertidal zone or from nearshore coastal habitats using subtidal dredges, diving, and plankton tows. Bering and Chukchi Sea samples were obtained from ship-based sampling in offshore waters (40–60 m water depth) using a van Veen grab and a plumb-staff beam trawl with 4 mm mesh. Samples were sieved on 1 and 0.5 mm mesh. Whenever possible, multiple specimens of each morphospecies were collected. Specimens were fixed in either 70 or 90% ethanol, which was replaced at least three times to prevent dilution, and preserved in 95% ethanol. Specimens were initially assigned to a family or genus using morphological criteria. Following DNA sequence analysis, specimens from each barcode cluster were morphologically identified using taxonomic keys for each region and, where possible, species identifications were verified by taxonomic specialists. Taxonomic assignments follow the World Register of Marine Species (WoRMS) [68]. In cases where a single species split into two or more distinct clusters, the Linnaean species name was retained and appended with interim identifiers (e.g. *Nereis pelagica* CMC01). Where the most precise identification was to a genus, similar interim names were applied (e.g. *Polycirrus* sp. CMC01). Morphological work is ongoing and taxonomic assignments will be updated as specimens are studied in more detail. Specimens will be deposited in the Canadian Museum of Nature and the Biodiversity Institute of Ontario. Collection data, photographs, specimen information, sequences, and trace files are available from the PONA project console on BOLD. Sequences have also been deposited in GenBank (accession nos. GU672096–GU672410, HM375142–HM375143, HM375485–HM375497, HM473288–HM473810, HM904906–HM904907, HM906747–HM906752, HQ023430–HQ023818, HQ023865–HQ024489).

DNA isolation, extraction, and amplification

A small piece of muscle tissue was lysed in 45 µl cetyltrimethylammonium bromide (CTAB) lysis buffer solution [69] plus 5 µl proteinase K. These samples were incubated at 56°C for 12–18 hours and DNA was then extracted using the manual protocol of Ivanova et al. [69] with a 3 µm glass fibre plate and re-suspended in 50 µl of ddH₂O. The target COI region was amplified using three primer sets (Table S1). PCR reactions were carried out in a 12.5 µl reaction volume containing 6.25 µl 10× reaction buffer, 0.625 µl MgCl₂ (50 mM), 0.125 µl of each primer (10 µM), 0.0625 µl dNTPs (10 mM), 0.06 µl Platinum Taq polymerase, and 2 µl of DNA template (10–50 ng). The thermocycling profile for all primers

![Figure 6. Faunal connectivity among sampling sites.](https://example.com/figure6.png)
To provide an overview of the similarity in species composition, matrices were constructed as Species, Sample, Abundance triplets. Sequences were manually edited using Sequerencer v.4.5 (Gene Codes Corporation, Ann Arbor, MI) and aligned by ClustalW in MEGA 4.0 [70]. Anomalous sequences that were not obvious pseudogenes (i.e., no frameshifts, full-length) or contaminants (following BLAST-type searches) were reanalyzed with alternate primers to test for multiple copies of COI. Twelve species had multiple indels in their COI sequence (Serpulidae: \( n = 9 \), Trissisic \( n = 2 \), and Syio setosa), but in all cases the reading frame was preserved suggesting the amplified COI fragments are functional copies.

Barcode clusters were initially assigned provisional species status using a 10 times average intrachanger divergence threshold [3,44]. To further assess provisional species boundaries among clusters that were less divergent than this threshold, statistical parsimony networks were analyzed using the Templeton, Cran- dall, Sins parsimony algorithm (TCS) [71,72]. This analysis constructs haplotype networks by inferring the most parsimonious branch connections at a 95% confidence level [72] and has been shown to accurately partition barcodes into species clusters [73]. Neighbour-joining (NJ) analysis using the K2P model [74] was conducted in MEGA 4.0 [70] to provide a graphical representation of the species tree, using one representative per MOTU. The K2P distance metric was chosen for consistency and comparability with other barcode studies and for its suitability when sequence divergences are low, such as for the construction of intra- and interspecies phylogenies [34,75]. Moreover, the NJ K2P method is a computationally efficient approach that can be performed on the BOLD platform, making it well suited for large-scale and exploratory identification of provisional species using DNA barcodes. For all subsequent data analyses, MOTUs were used as proxies for species and genetic distances were calculated using the BOLD “Distance Summary” and “Nearest Neighbour Summary” tools [33].

Regional species richness and similarity indices
To examine differences in molecular and morphological richness, MOTUs were compared to identified species in the cases where morphological designations were available. Individual-based rarefaction curves (\( S_{obs} \), Mao Tau) were generated separately for identified species (\( n = 1343 \)) and for MOTUs (\( n = 1876 \)) using the software EstimateS v.8.2.0 [76] with 50 randomizations and sampling without replacement. Diversity Data matrices were constructed as Species, Sample, Abundance triplets. To provide an overview of the similarity in species composition between collection sites, Chao’s Abundance-based Sørensen similarity index [77,78] was calculated. This shared species estimator modifies the classic Sørensen similarity index for presence-absence data by accounting for unequal or incomplete sampling of a region and is particularly useful when there are many rare species in a dataset [70]. Shared Species matrices were constructed in EstimateS [76] as Species, Sample, Abundance triplets with seven collection sites (samples) defined. Maps were obtained from SimpleMapp [79].

Statistical analysis
To determine whether the number of individuals analyzed affected the level of variation within genetic clusters, a linear regression was performed. To test the assumptions of homoscedastic residual variance, a Spearman rank correlation was run on the absolute value of residual variation in K2P distance and the number of individuals sampled. A Shapiro-Wilk test was used to test if residuals were normally distributed, but no significant deviations were detected.

Supporting Information
Figure S1 Neighbour-joining tree for 333 provisional polychaete species. One specimen per MOTU is shown with abundances indicated in brackets. Collection locations are indicated by pie graphs. Vertical bars indicate identified species whose members fall into two or more MOTUs. (PDF)

Table S1 List of primers used in this study. (PDF)

Table S2 Polychaete family richness and abundance by sampling site. (PDF)

Acknowledgments
We thank Dieter Fiege, Myriam Schüller, Max Hoberg, Ruth Barnich, and Torkild Bakken for aiding in specimen identification. We are also grateful to Katrina Iken for collection of Bering Sea samples, to Erin Costorphine for assistance with fieldwork in Bamfield, Churchill, and St. Andrews, and to Joey Angnaton, the crew of the What’s Happening, Mallory Carpenter, and Dorothy Angnaton for their support and expertise in the field at Labrador. We thank Sarah Adamowicz, Mark Sherrard, John Wilson, and Alex Smith for helpful comments on earlier versions of this manuscript.

Author Contributions
Conceived and designed the experiments: CMC PDNH. Performed the experiments: CMC. Analyzed the data: CMC. Contributed reagents/materials/analysis tools: CMC SMH TMB TAM PDNH. Wrote the paper: CMC SMH TMB TAM PDNH. Collected/identified specimens: CMC SMH TMB TAM PDNH.

References
1. Westheide W, Schmidt H (2003) Cosmopolitan versus cryptic meiofaunal polychaete species: An approach to a molecular taxonomy. Helgol Mar Res 57: 1–6.

2. Worheide G, Solé-Cava AM, Hooper JNA (2005) Biodiversity, molecular ecology and phylogeny of marine sponges: Patterns, implications and outlooks. Integr Comp Biol 45: 377–385.

3. Witt JDS, Threlfall DL, Hebert PDN (2006) DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: Implications for desert spring conservation. Mol Ecol 15: 3073–3082.

4. Palumbi SR (1994) Generic divergence, reproductive isolation, and marine speciation. Annu Rev Ecol Syst 25: 547–572.

5. Radulovici AE, Archambault P, Dufrasne F (2010) DNA barcodes for marine biodiversity: Moving fast forward? Diversity 3: 450–472.

6. Knowlton N (1993) Sibling species in the sea. Annu Rev Ecol Syst 24: 189–216.

7. Knowlton N (2000) Molecular genetic analyses of species boundaries in the sea. Hydrobiol 420: 73–90.

8. Quinon PA, Smelgrov PVR (2005) Polychaete assemblages of a sub-arctic Newfoundland fjord: Habitat, distribution, and identification. Polar Biol 28: 495–505.

9. Hutchings P, Fauchald C (2000) Class Polychaeta. Definition and general description. In: Beesley PL, Ross GJB, Glasby CJ, eds. Polychaetes and allies: The southern synthesis. Melbourne: CSIRO Publishing. pp 1–495.

10. Jirkov IA (2001) Polychaeta of the Arctic Ocean. Moscow: Yamus-K. 632 p.

11. Rose-GW, Fleig F (2001) Polychaetes. London: Oxford University Press. pp 153–495.

12. Fauchald K (1984) Polychaete distribution patterns, or: Can animals with Palaeozoic cousins show large-scale geographical patterns? In: Hutchings PA,
74. Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16: 111–120.
75. Nei M, Kumar S (2000) Molecular evolution and phylogenetics. New York: Oxford University Press, 333 p.
76. Colwell RK (2009) EstimateS: Statistical estimation of species richness and shared species from samples. Version 8.2. Available: http://purl.oclc.org/estimates. Accessed 2010 Nov 10.
77. Sørensen TA (1948) A method of establishing groups of equal amplitude in plant sociology based on similarity of species content, and its application to analyses of the vegetation on Danish commons. K Dan Vidensk Selsk Biol Skr 5: 1–34.
78. Chao A, Chazdon RL, Colwell RK, Shen TJ (2005) A new statistical approach for assessing similarity of species composition with incidence and abundance data. Ecol Lett 8: 148–159.
79. Shorthouse DP (2010) SimpleMappr, a web-enabled tool to produce publication-quality point maps. Available: http://www.simplemappr.net. Accessed 2010 Oct 14.