Journal of Fish Biology (2008) **72,** 464–471
doi:10.1111/j.1095-8649.2007.01745.x, available online at http://www.blackwell-synergy.com

DNA barcoding for the identification of smoked fish products

P. J. Smith*†, S. M. McVeagh* and D. Steinke‡

*National Institute of Water and Atmospheric Research Ltd, Private Bag 14 901, Wellington, New Zealand and ‡Canadian Centre for DNA Barcoding, Biodiversity Institute of Ontario, University of Guelph, 579 Gordon Street, Guelph, Ontario, N1G 2W1 Canada

(Received 11 July 2007, Accepted 17 October 2007)

DNA barcoding was applied to the identification of smoked products from fish in 10 families in four orders and allowed identification to the species level, even among closely related species in the same genus. Barcoding is likely to become a standard tool for identification of fish specimens and products.

Key words: DNA barcoding; smoked fish product; species identification.

Identification of fish fillets usually requires the application of a molecular tool because most of the morphological traits used in species identification are removed during the filleting process. Early molecular identification methods relied on protein electrophoresis that became recognized as an official method for fish fillet identification (AOAC, 1990; Tenge *et al.*, 1993; Yearsley *et al.*, 1999). Proteins are stable in fresh and frozen product, but are denatured and damaged by heat and salt processing, and are generally not suitable markers for species identification in smoked and canned fish products (Sotelo *et al.*, 1992; Unlusayan *et al.*, 2001). Furthermore, closely related fish species may share protein profiles, precluding specific identification of product (Bartlett & Davidson, 1991; Smith *et al.*, 1996). The rapid developments in molecular biology have provided a range of tools for fish specimen and product identification. Most methods are based on amplification of a specific gene region and restriction enzyme digests of the amplified products and have been used to identify and distinguish closely related fish species. Examples include the cytochrome *b* gene in European flatfish (Cespedes *et al.*, 1998), the 16S rRNA in salmonids (Carrera *et al.*, 1999), the control region in swordfish (Hsieh *et al.*, 2004) and the *p53* nuclear gene in salmonids (Carrera *et al.*, 2000). However, the methods

†Author to whom correspondence should be addressed. Tel.: +64 4 3860300; fax: +64 4 3860574; email: p.smith@niwa.co.nz

are often order or family specific and not universally applicable (Hsieh *et al.*, 2004). Multiplex assays and probe methods developed for specific groups are likewise not universally applicable (Itoi *et al.*, 2005; Shivji *et al.*, 2005), although the probe methods are applicable to short DNA sequences recovered from heat-processed samples (Itoi *et al.*, 2005).

A global DNA-based barcode identification system is being developed for all animal species and when coupled with high throughput sequencing methods potentially provides a simple, universal tool for the identification of fish species and products. The barcode system is based on DNA diversity in a single gene region (a section of the mitochondrial DNA cytochrome *c* oxidase I gene, COI); suspect specimens and products are identified by comparing their DNA barcode sequences against a DNA database of COI sequences derived from reference specimens. Hebert *et al.* (2003*a*, *b*) have demonstrated that the COI region is appropriate for discriminating between closely related species across diverse animal phyla, and this has been verified in marine (Ward *et al.*, 2005) and freshwater fishes (Steinke *et al.*, 2005; Ward *et al.*, 2005). DNA barcodes have been obtained for over 4000 species of fish, including 400 species from the New Zealand Exclusive Economic Zone, and the COI sequences deposited in the Barcode of Life Database (BOLD, www.boldsystems.org) (Ratnasingham & Hebert, 2007) and additional fish COI sequences are available in GenBank. The present study was undertaken to identify species in products labelled smoked eel in New Zealand, where three species of freshwater eel (*Anguilla*) and at least four species of marine eel (*Conger* and *Bassanago*) are fished commercially. The barcode method was further evaluated for the identification of smoked fillets from species in 10 families in four different orders: Salmoniformes, Anguilliformes, Gadiformes and Perciformes.

Samples of smoked eel were provided by New Zealand Ministry of Fisheries compliance staff. Additional samples of smoked fish were purchased at supermarkets (Table I). Fillet samples taken from kahawai *Arripis trutta* (Forster 1801) identified by a taxonomist were hot smoked in a domestic smoker. Small sub-samples of muscle tissue (200–500 mg) were taken from a selection of the smoked fish products, and total genomic DNA was extracted by homogenization and digestion with proteinase-K at 55° C for 4 h, followed by a standard phenol–chloroform–ethanol procedure (Taggart *et al.*, 1992).

Approximately 600 base pairs (bp) of the 5′ region of the COI gene were amplified using the primer pair FishF2 and FishR2 (Ward *et al.*, 2005) for each muscle tissue sub-sample. Amplifications were carried out using an initial denaturation of 94° C for 1 min; 35 cycles of 94° C for 60 s, 54° C for 90 s and 72° C for 60 s, followed by an extension at 72° C for 5 min, using a Cetus 9600 DNA thermocycler (PerkinElmer Corporation, CT, U.S.A.). Polymerase chain reaction (PCR) products were purified using the QIAquick gel extraction kit (Qiagen Pty Ltd, Doncaster, Victoria, Australia). Sequences were determined using the ABI Taq DyeDeoxy™ Terminator Cycle Sequencing Kit according to the manufacture's directions (Applied Biosystems Inc., Foster City, CA, U.S.A.) and run on an ABI 3730 autosequencer. DNA sequences were edited in CHROMAS (Technelysium, Queensland, Australia) and aligned in CLUSTAL in MEGA version 3 (Kumar *et al.*, 2004) and stored in BIOEDIT (Hall, 1999).

TABLE I. Smoked fish product, possible New Zealand species and barcode identities

Product name	Likely species (family, order)	Per cent identity (bases match)
Eel	*Anguilla australis* shortfin eel	94 (582/621)
	Anguilla dieffenbachii longfin eel	99–100 (620/621; 621/621)
	Anguilla reinhardtii speckled longfin eel (Anguillidae, Anguilliformes)	93 (577/621)
	Bassanago bulbiceps (Whitley 1948) swollen headed conger	83–86 (515/621; 534/621)
	Bassanago hirsutus Castle 1960 hairy conger	83 (515/621)
	Conger verreauxi Kaup 1856 southern conger	83 (515/621)
	Conger wilsoni (Bloch & Schneider 1801) northern conger (Congridae, Anguilliformes)	83 (515/621)
Salmon	*Oncorhynchus tshawytscha* (Walbaum 1792) Chinook salmon (Salmonidae, Salmoniformes)	99–100 (580/581; 581/581)
Hoki	*Macruronus novaezelandiae* (Hector 1871) hoki (Macruronidae, Gadiformes)	99 (492/495)
Groper	*Polyprion oxygeneios* (Schneider & Forster 1801) groper and hapuku	99 (521/523)
	Polyprion americanus (Bloch & Schneider 1801) bass (Polyprionidae Perciformes)	95 (503/524)
Trevally	*Pseudocaranx dentex* (Bloch & Schneider 1801) trevally (Carangidae, Perciformes)	100 (582/582)
Mackerel *Trachurus novaezelandiae*	*Trachurus declivis* jack mackerel	100 (578/578)
	T. novaezelandiae horse mackerel	99 (574/578)
	Trachurus murphyi Murphy's mackerel (Carangidae Perciformes)	98 (569/578)
Kingfish	*Seriola lalandi* Valenciennes 1833 kingfish (Carangidae Perciformes)	100 (581/581)
Snapper	*Pagrus auratus* (Forster 1801) snapper (Sparidae Perciformes)	99 (553/555)
Kahawai	*Arripis trutta* kahawai (Arripididae Perciformes)	100 (509/509)
Blue cod	*Parapercis colias* (Forster & Schneider 1801) blue cod (Pinguipedidae Perciformes)	100 (581/581)

TABLE I. Continued

Product name	Likely species (family, order)	Per cent identity (bases match)
Tuna *Thunnus alalunga*	*Thunnus alalunga* (Bonnaterre 1788) albacore tuna	99–100 (538/540; 540/540)
	Thunnus albacares (Bonnaterre 1788) yellowfin tuna	98 (532/540)
	Thunnus obesus (Lowe 1839) bigeye tuna	97 (520/530)
	Thunnus maccoyii (Castelnau 1872) Southern bluefin tuna	98 (532/540)
	Thunnus orientalis (Temmink & Schlegel 1844) Pacific northern bluefin tuna	98 (535/540)
	Katsuwonus pelamis (Linnaeus 1758) skipjack tuna (Scombridae, Perciformes)	89 (469/526)
Warehou	*Seriolella brama* (Günther 1860) blue warehou	100 (581/581)
	Seriolella caerulea Guichenot, 1848 white warehou	95 (553/581)
	Seriolella punctata (Forster 1801) silver warehou (Centrolophidae, Perciformes)	93 (546/581)

Sequences from the smoked eel products were aligned against GenBank COI entries for the longfin eel *Anguilla dieffenbachii* Gray 1842 (accession numbers AP007240 and NC006538), the shortfin eel *Anguilla australis* Richardson 1841 (NC006532, NC006534 and AP007234–35), the speckled longfin eel *Anguilla reinhardtii* Steindachner 1867 (NC006546 and AP007248), New Zealand *Conger* (EU182964–66) and New Zealand *Bassanago* (EU182960–63 and EU182967) and unpublished sequences held at the National Institute of Water and Atmospheric Research for *A. dieffenbachii*. Sequence divergences were calculated using the Kimura two-parameter (K2P) distance model (Kimura, 1981). A neighbour-joining (NJ) tree of K2P distances was created to provide a graphic representation of divergence among the eel species and the suspect-smoked eel product, using the software MEGA3 (Kumar *et al.*, 2004); bootstrapping was performed with 1000 replications, and the *Bassanago* sequences used to root the tree.

COI sequences from the smoked blue cod, groper, hoki, kahawai, kingfish, mackerel, salmon, snapper, trevally, tuna and warehou fillets were matched against COI sequences taken from reference specimens held in BOLD and GenBank, and sequence identities >98% listed (Table I). Sequences from the smoked mackerel product were aligned against GenBank entries for jack mackerel *Trachurus declivis* (Jenyns 1841) (EU182969–72), horse mackerel *Trachurus novaezelandiae* Richardson 1843 (EU182959, EU182975–78) and Murphy's

mackerel or Inca scad *Trachurus murphyi* Nichols 1920 (EU182973–74) from New Zealand. A NJ tree of K2P distances was created in MEGA3 (Kumar *et al.*, 2004), and the *T. murphyi* sequences used to root the tree.

Sequences 520–621 bp in length were obtained from all smoked fish products with the exception of one (of three) smoked fillets labelled as hoki. The heating and salting (=drying) techniques used to produce smoked fish do not appear to denature the DNA to an extent that eliminates recovery of moderate length sequences from processed fillets. Previous studies of raw and cold-smoked salmon (*Oncorhynchus* and *Salmo*) had recovered PCR products *c.* 530 bp in the *p53* nuclear gene (Carrera *et al.*, 2000) and *c.* 950 bp in the mitochondrial 16S gene (Carrera *et al.*, 1999), but the restriction enzyme digest approach used with these identification methods was only applicable to a small range of species, and it was recognized that more than one genetic marker would be required to demonstrate substitution in a range of species using the restriction fragment length polymorphisms (RFLP) approach (Carrera *et al.*, 2000).

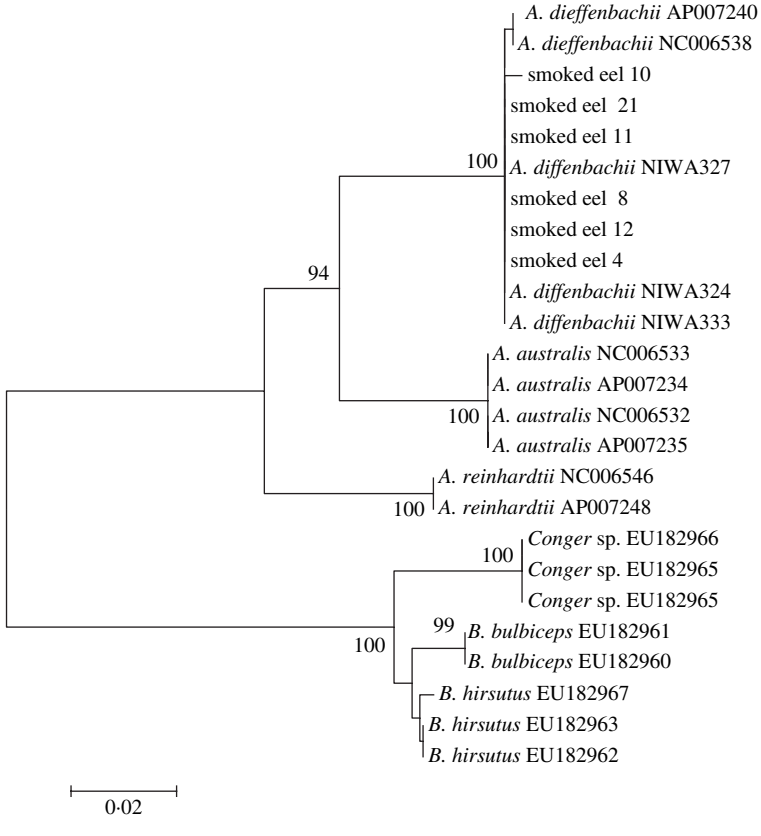

FIG. 1. Relationships of cytochrome *c* oxidase I gene (COI) sequences from smoked eel products and reference eel species. COI sequence data are as described in the text with GenBank accession numbers. Numbers at nodes are bootstrap percentages (>80%) after 1000 replicates, based on distance; scale bar represents an interval of the Kimura two-parameter (K2P) model.

One pair of universal primers successfully amplified the COI region in smoked fillets from a range of teleosts in the orders Anguilliformes, Salmoniformes, Gadiformes and Perciformes. The smoked fillets could be identified to the species level with 99–100% identity matches against sequences held in BOLD and GenBank, even among closely related species in the same genus in New Zealand waters (Table I: *Thunnus*, *Seriolella*, *Polyprion* and *Trachurus*).

Unambiguously aligned sequences were obtained for 621 bp of COI for the *Anguilla*, *Conger* and *Bassanago* data set. COI sequences from the New Zealand smoked eel products aligned with the COI sequences for the longfin eel *A. dieffenbachii* (99–100% identity) and not shortfin (94% identity) or speckled longfin eels (93%), *A. australis* and *A. reinhardtii*, respectively (Table I and Fig. 1). Much lower identities were found between the smoked eel product and the marine *Conger* and *Bassanago* eels (81–85% identities). It was concluded that the smoked eel product was *A. dieffenbachii*.

Unambiguously aligned sequences were obtained for 520–600 bp of COI sequences from the smoked blue cod, groper, hoki, kahawai, kingfish, mackerel, salmon, snapper, trevally, tuna and warehou fillet samples. All of the smoked fish products purchased in retail outlets were labelled with the correct common name, but smoked product labelled as mackerel *T. novaezelandiae* was mackerel *T. declivis* (Table I and Fig. 2). The two species of *Trachurus* are difficult to discriminate, the key morphological character being the relative length of the dorsal accessory lateral line (Paulin *et al.*, 2001), and specimens are only readily distinguished by fisheries biologists and taxonomists.

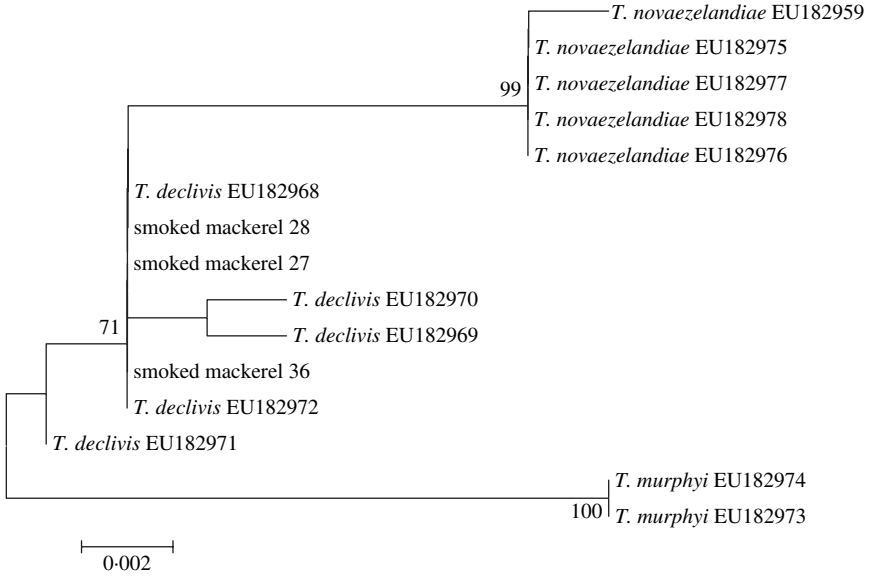

FIG. 2. Relationships of cytochrome *c* oxidase I gene (COI) sequences from smoked mackerel products and reference *Trachurus* species from New Zealand. COI sequence data are as described in the text with GenBank accession numbers. Numbers at nodes are bootstrap percentages (>70%) after 1000 replicates, based on distance; scale bar represents an interval of the Kimura two-parameter (K2P) model.

Unlike the RFLP (Hsieh *et al.*, 2004), multiplex assay (Shivji *et al.*, 2005), amplified fragment length polymorphism (AFLP) (Maldini *et al.*, 2006) and single nucleotide polymorphism (Itoi *et al.*, 2005) methods developed for specific fish fillet and product identifications, the barcoding method is universal, with one set of primers applicable to a wide range of fishes. AFLP markers have the potential to discriminate among regional stocks but require an increasing number of primer combinations as taxonomic distance among species is reduced; and the co-migration of non-homologous fragments can further reduce discrimination (Maldini *et al.*, 2006). As BOLD expands and high throughput sequencing tools become more accessible, barcoding is likely to become a standard tool for the identification of fish specimens and products, including smoked fillets.

We thank N. Bagley and P. McMillan, NIWA, for specimen collection and identification of reference specimens. C. Roberts and A. Stewart, Museum of New Zealand Te Papa Tongarewa, provided tissue samples for some reference specimens of *Bassanago*, *Conger* and *Polyprion* as part of Te Papa's Collection Development Programme. D.S. was supported through funding to the Canadian Barcode of Life Network from Genome Canada through the Ontario Genomics Institute. This project was supported by funding from the New Zealand Ministry of Fisheries, Project GDB200601.

References

AOAC (1990). *Official Methods of Analysis*, 15th edn.Washington, DC: Association of Official Analytical Chemists.

Bartlett, S. E. & Davidson, W. S. (1991). Identification of *Thunnus* tuna species by the polymerase chain reaction and direct sequence analysis of their mitochondrial cytochrome b genes. *Canadian Journal of Fisheries and Aquatic Sciences* **48,** 309–317.

Carrera, E., Garcia, T., Céspedes, A., González, I., Fernández, A., Hernández, P. E. & Martin, R. (1999). Salmon and trout analysis by PCR-RFLP for identity authentication. *Journal of Food Science* **64,** 410–413. doi: 10.1111/j.1365-2621.1999.tb15053.x

Carrera, E., García, T., Céspedes, A., González, I., Fernández, A., Asensio, L. M., Hernández, P. E. & Martín, R. (2000). Identification of smoked Atlantic salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*) using PCR-restriction fragment length polymorphism of the p53 gene. *Journal of AOAC International* **83,** 341–346.

Cespedes, A., Garcia, T., Carrera, E., Gonzalez, I., Sanz, B., Hernaz, P. E. & Martin, R. (1998). Identification of flatfish species using polymerase chain reaction (PCR) amplification and restriction analysis of the cytochrome b gene. *Journal of Food Science* **63,** 206–209. doi: 10.1111/j.1365-2621.1998.tb15710.x

Hall, T. (1999). Bioedit: a user friendly biological sequence alignment editor and analysis programme for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41,** 95–98.

Hebert, P., Cywinska, A., Ball, S. & deWaard, J. (2003*a*). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B* **270,** 313–322.

Hebert, P. D. N., Ratnasingham, S. & deWaard, J. R. (2003*b*). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London B* **270,** S96–S99.

Hsieh, H. S., Chai, T., Cheng, C. A., Hsieh, Y. W. & Hwang, D. F. (2004). Application of DNA technique for identifying the species of different processed products of swordfish meat. *Journal of Food Science* **69,** FCT1–FCT6. doi: 10.1111/j.1365-2621.2004.tb17847.x

Itoi, S., Nakaya, M., Kaneko, G., Kondo, H., Sezaki, K. & Watabe, S. (2005). Rapid identification of eels *Anguilla japonica* and *Anguilla anguilla* by polymerase chain reaction with single nucleotide polymorphism-based specific probes. *Fisheries Science* **71,** 1356–1364. doi: 10.1111/j.1444-2906.2005.01102.x

Kimura, M. (1981). Estimation of evolutionary distances between homologous nucleotide sequences. *Proceedings National Academy of Sciences* **78,** 454–458.

Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* **5,** 150–163.

Maldini, M., Nonnis Marzano, F., Fortes, G., Papa, R. & Gandolfi, G. (2006). Fish and seafood traceability based on AFLP markers: elaboration of a species database. *Aquaculture* **261,** 487–494.

Paulin, C., Stewart, A., Roberts, C. & McMillan, P. (2001). *New Zealand Fishes: a Complete Guide*. Wellington: Te Papa Press.

Ratnasingham, S. & Hebert, P. D. (2007). The barcode of life database. *Molecular Ecology Notes* **7,** 355–364. doi: 10.1111/j.1471-8286.2007.01678.x

Shivji, M. S., Chapman, D. D., Pikitch, E. K. & Raymond, P. W. (2005). Genetic profiling reveals illegal international trade in fins of the great white shark, *Carcharodon carcharias*. *Conservation Genetics* **6,** 1035–1036.

Smith, P. J., Roberts, C. D., McVeagh, S. M. & Benson, P. G. (1996). Genetic evidence for two species of tarakihi (Teleostei: Cheilodactylidae: Nemadactylus) in New Zealand waters. *New Zealand Journal of Marine and Freshwater Research* **30,** 209–220.

Sotelo, C. G., Piñeiro, C., Gallardo, J. M. & Pérez-Martin, R. I. (1992). Identification of fish species in smoked fish products by electrophoresis and isoelectric focusing. *Zeitschrift für Lebensmitteluntersuchung und -Forschung A* **195,** 224–227.

Steinke, D., Vences, M., Salzburger, W. & Meyer, A. (2005). TaxI: a software tool for DNA barcoding using distance methods. *Philosophical Transactions of the Royal Society, London, Series B* **360,** 1847–1857.

Taggart, J. B., Hynes, R. A., Prodohl, P. A. & Ferguson, A. (1992). A simplified protocol for routine total DNA isolation from salmonid fishes. *Journal of Fish Biology* **40,** 963–965. doi: 10.1111/j.1095-8649.1992.tb02641.x

Unlusayan, M., Kaleli, S. & Gulyavuz, H. (2001). The determination of flesh productivity and protein components of some fish species after hot smoking. *Journal of the Science of Food and Agriculture* **81,** 661–664.

Ward, R. D., Zemlak, T. S., Ines, B. H., Last, P. R. & Hebert, P. D. N. (2005). DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society, London, Series B* **360,** 1847–1857.

Yearsley, G. K., Last, P. R. & Ward, R. D. (1999). *Australian Seafood Guide: an Identification Guide to Domestic Species*. Hobart, Tasmania: CSIRO Marine Research.

Electronic Reference

Tenge, B., Dang, N. L., Fry, F., Savary, W., Rogers, P., Barnett, J., Hill, W., Rippey, S., Wiskerchen, J. & Wekell, M. (1993). The regulatory fish encyclopedia: an internet-based compilation of photographic, textural and laboratory aid in species identification of selected fish species. U.S. Food and Drug Administration. Available at http://vm.cfscan.fda.gov/~frf/rfe0.html

