BioChip J. (2011) 5(1): 72-77
DOI 10.1007/s13206-011-5111-6

Original Research

DNA chip for species identification of Korean freshwater fish: A case study

Sungmin Kim[1], Hyeyoung Koo[2], Ji-Hoon Kim[3], Jin-Wook Jung[3], Seung Yong Hwang[3] & Won Kim[1,4]

Received: 15 October 2010 / Accepted: 22 December 2010 / Published online: 20 March 2011

Abstract DNA barcoding is a molecular diagnostic method for species identification that uses a single standardized DNA fragment. Remarkably, mitochondrial cytochrome *c* oxidase subunit I (*COI*) gene in animal species can be used for this purpose. For molecular identification, there are several approaches available based on varying properties such as sequence similarity, length of PCR products, and hybridization. We previously developed web-based Molecular Identification System for Fish (MISF), including 53 Korean freshwater fish species, based on a profile hidden Markov model and sequence similarity. In this study, we developed a DNA chip arrayed with 16 oligonucleotide probes to identify 11 selected species of Korean freshwater fish. The *COI* gene was quite suitable for designing species-specific oligonucleotide probes and a DNA chip arrayed with these probes showed high resolution for species identification. Therefore, the DNA chip using the *COI* gene can be further developed for different purposes by optimal species selection in biodiversity studies and environmental monitoring.

Keywords: DNA chip, Freshwater fish, Species identification, DNA barcoding, *COI*

[1]Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 151-747, Korea
[2]Department of Biological Science, Sangji University, Wonju 220-702, Korea
[3]Department of Biochemistry, Hanyang University & GenoCheck Co., Ltd., Sangnok-gu, Ansan, Gyeonggi-do 426-791, Korea
[4]School of Biological Sciences, Seoul National University, Seoul 151-747, Korea
Correspondence and requests for materials should be addressed to W. Kim (✉wonkim@plaza.snu.ac.kr)

Introduction

DNA-based identification methods are a necessity in biodiversity studies[1]. Although not completely error-free, DNA barcoding techniques have recently shown promise in biodiversity studies. DNA barcoding is a molecular diagnostic method for rapid and accurate species identification of species[2,3]. This method not only facilitates fast and accurate identification, but also allows for effective molecular identification, regardless of life stage (such as egg, larva, and adult) or specimen quality (such as a broken or imperfect specimen with missing parts) and does not require previous taxonomic knowledge[4-6]. Several studies have confirmed that a 648-base pair (bp) segment in the 5′ region of mitochondrial cytochrome c oxidase subunit I (*COI*) gene can be used as a DNA barcode to identify most animal species[7-10]. Currently, DNA barcode libraries are available in large-scale biodiversity databases such as the Barcode of Life Data Systems (BOLD; http://www.boldsystems.org)[11] and the Korea Barcode of Life (KBOL; http://koreabarcode.org).

Several molecular identification approaches use a different property of DNA barcode such as match of sequence similarity, different lengths of PCR products[12], or species-specific hybridization[13]. For comparing of sequence similarity, many sophisticated methods and algorithms, such as the basic local alignment search tool (BLAST)[14], TaxI software[15], Taxon ID tree[11], DNA-BAR[16], and character-based search[17] have been used. We used a profile hidden Markov model and the property of sequence similarity to develop a web-based Molecular Identification System for Fish (MISF) to identify 53 Korean freshwater fish species, 233 oth-

er freshwater fish species and 1139 saltwater fish species[18].

DNA chip is a new powerful and innovative tool that uses microarray technology[19,20] and the degree of hybridization of complementary DNA sequences. Recently, DNA chip-based methods have achieved high resolution for identification of species of several organisms[21,22]. In this study, we examined the utility of the *COI* gene as an oligonucleotide probe for use in a DNA chip array[23] and exploited the usefulness of DNA chip for species-level identification of Korean freshwater fish. Thus, we present a case study of application of a DNA chip method for the identification of 11 Korean freshwater fish species.

Results and Discussion

Selection of target species-specific oligonucleotide

We designed 33 potential probes of 11 target species on the basis of 3 probes for each species. The probe is a 23-nucleotide long DNA segment (Figure 1A) with the specifications rather tolerant for GC content and melting temperature (T_m), as recommended by Pfunder *et al.*, 2004[20]. We constructed a temporal DNA chip arrayed using these 33 probes. Using this chip, the best candidate probes of each species were selected by hybridization to 325 *COI* barcodes consisting of target (n=11) and non-target (n=20) species. We selected 2 probes from each of 5 species (*Zacco platypus, Rhinogobius giurinus, Rhinogobius brunneus, Opsariichthys uncirostris amurensis*, and *Acheilognathus rhombeus*) and 1 probe in the remaining 6 species (*Carassius auratus, Acheilognathus yamatsutae, Zacco koreanus, Rhodeus notatus, Rhodeus uyekii*, and *Acanthorhodeus gracilis*). A total of 16 probes were selected. However, the length of 1 probe, R.bru_1, was decreased to 22-nucleotides to attain a similar T_m value to the other probes. Probe melting temperatures were between 51.7°C and 58.8°C with a GC content of 39.1-65.0% (Table 1).

Discrimination of target species

The degree of hybridization signal on the DNA chip was our most significant concern in this study. The DNA chip experiment showed that most target hybridization true positive signals (average 9009.3 arbitrary units) were distinctly stronger than other signals (average 87.1 arbitrary units). High standard deviations of true positive signals and cross-hybridization signals were observed. Due to variation in individual hybridization signals, a standard signal threshold was established. We defined 4000 arbitrary units as a threshold after comparison with signals of non-target species to

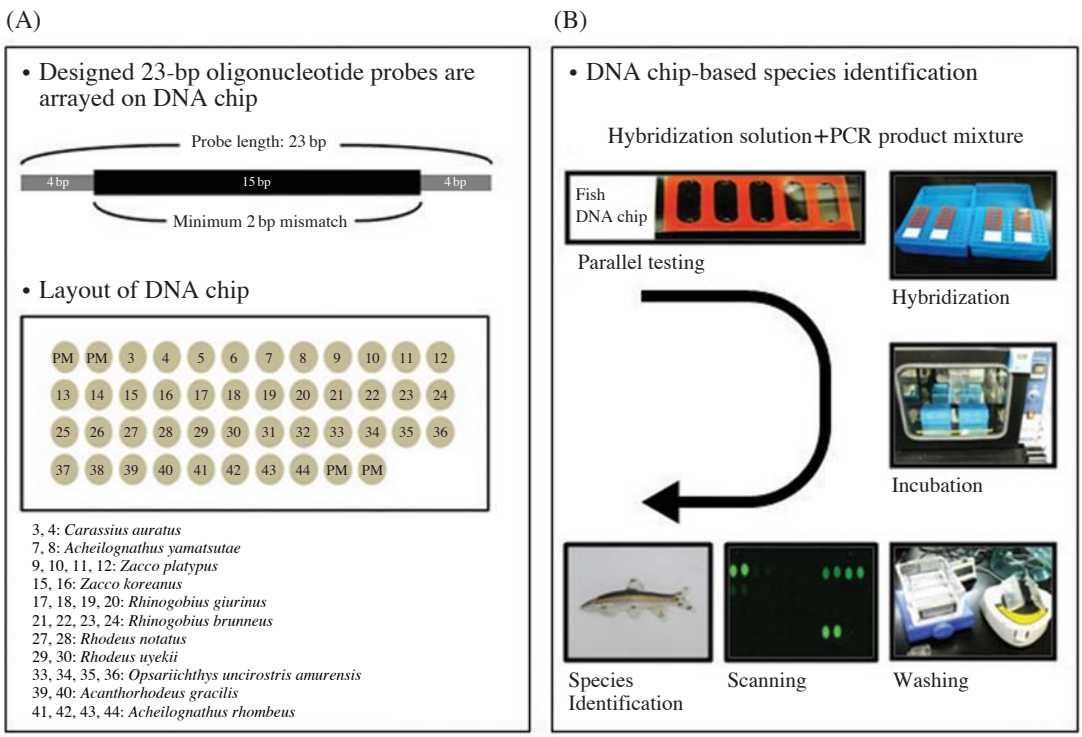

Figure 1. (A) Oligonucleotide design and DNA chip layout. (B) Procedure for species identification with DNA chip.

Table 1. Species-specific oligonucleotide probes for species identification.

Species	Probe	Specific sequence (5′-to 3′)	GC (%)	Tm (°C)	Length (bp)
	PM*	CATCCCCCTGGGACTGGAGT	65.0	58.0	20
Carassius auratus	C.aur_1	TAACCGCCGTCCTCCTTCTCCTA	56.5	58.8	23
Acheilognathus yamatsutae	A.yam_2	CTTCTATCGCTACCCGTTCTGGC	56.5	58.8	23
Zacco platypus	Z.pla_1	CTTCTCCTGTCCTTACCCGTACT	52.2	57.1	23
	Z.pla_2	CACCCTTATTCGTCTGAGCCGTA	52.2	57.1	23
Zacco koreanus	Z.kor_3	CCTTTTTGTATGAGCTGTACTTG	39.1	51.7	23
Rhinogobius giurinus	R.giu_2	CAGACCGCAACTTAAACACAACC	47.8	55.3	23
	R.giu_3	GAGCCGTTCTAATTACAGCTGTC	47.8	55.3	23
Rhinogobius brunneus	R.bru_1	CTTTCCCTTCCCGTTCTTGCCG	59.1	58.6	22
	R.bru_3_AS	GGCTAAGGGCCGTACCTACTATT	52.2	57.1	23
Rhodeus notatus	R.not_3_AS	TGAGCCTGGCTGGCTAAGTTCAG	56.5	58.8	23
Rhodeus uyekii	R.uye_1	GCTATCTTTACCAGTCTTGGCCG	52.2	57.1	23
Opsariichthys uncirostris amurensis	O.unc.amu_1	CCGGGATTACGATGCTCCTTACG	56.5	58.8	23
	O.unc.amu_2	CTATCCCTACCCGTATTGGCTGC	56.5	58.8	23
Acanthorhodeus gracilis	A.gra_2	GTACTTCTCCTTTTATCTCTACC	39.1	51.7	23
Acheilognathus rhombeus	P.rho_2	GTACTTCTACTGCTATCACTACC	43.5	53.5	23
	P.rho_3	AACACCTTTATTCGTATGAGCTG	39.1	51.7	23

PM*: Position Marker, non-homologous Cy3-labeled probe

Table S1. Magnitude of target fluorescence signals in true-positives.

Target species	Probe name	No. of samples	Mean absolute fluorescence signal in arbitrary units	Standard deviation
Carassius auratus	C.aur_1	3	9753.7	±2400.3
Acheilognathus yamatsutae	A.yam_2	4	8157.4	±2061.0
Zacco platypus	Z.pla_1	3	4400.7	±312.2
	Z.pla_2	3	5210.2	±352.5
Zacco koreanus	Z.kor_3	4	8436.6	±2562.6
Rhinogobius giurinus	R.giu_2	4	12324.8	±3238.9
	R.giu_3	4	8009.0	±823.0
Rhinogobius brunneus	R.bru_1	4	10923.0	±2396.7
	R.bru_3_AS	4	8865.3	±4448.6
Rhodeus notatus	R.not_3_AS	3	12191.7	±3858.6
Rhodeus uyekii	R.uye_1	2	9301.3	±102.2
Opsariichthys uncirostris amurensis	O.unc.amu_1	4	8556.4	±1030.0
	O.unc.amu_2	4	16486.5	±4527.8
Acanthorhodeus gracilis	A.gra_2	3	6069.3	±1094.7
Acheilognathus rhombeus	P.rho_2	3	6779.5	±352.2
	P.rho_3	3	8683.0	±1110.7

maximize the number of true positive and negative values. The threshold value suggests that target hybridization true positive signals range from 4400.7 to 16486.5 arbitrary units (Table S1), and true negative signals range from −103.5 to 3782.3 arbitrary units. In particular, O.unc.amu_2, 1 of 2 *O. uncirostris* amurensis target probes, was erroneously cross-hybridized with *R. giurinus*, and its signal value was 4619.0. However, because O.unc.amu_2 showed a relatively high signal intensity of 16486.5 in the species sample hybridization and 2 *R. giurinus* probes also showed high signal intensities of over 8000 arbitrary units (Figure 2), we were able to eliminate this data from the true positive category. We confirmed that DNA chip-based diagnosis performs well on the species level and with high sensitivity (99%) and specificity (100%) from 1424 hybridizations (89 sample product of 31 species × 16 DNA chip probes) (Figure 2).

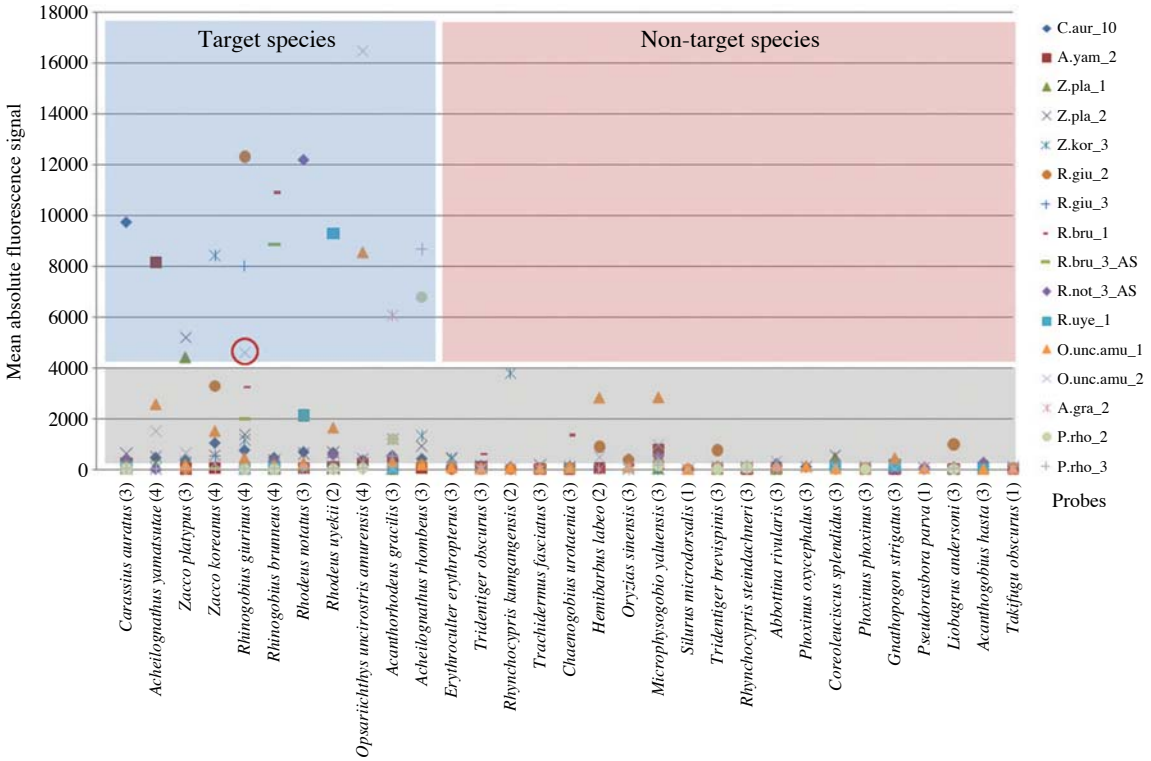

Figure 2. Mean absolute fluorescence signals of target species are entirely observed in the blue box (true positive), while those of non-target species are not detected and shown in the red box (true negative) above the threshold, 4000 (grey box). Only one signal is observed as a false positive value in target hybridizations (red circle). The number of hybridizations is given in brackets.

DNA chip technology is an attractive method for species identification because it is fast, highly accurate, cost-effective, and requires little training[13]. DNA chip probes are not only sensitive to the target species but also independent to all non-target species. Additionally, multiple detections are possible. However, probes design should be based on geographical genetic variation for large-scale monitoring survey because it can be limited by unknown haplotypes[24]. The reproducibility, accuracy, and signal intensity threshold of all probes must be evaluated to maximize DNA performance[25].

Conclusions

DNA chip was used to array 16 oligonucleotide probes of 11 target species to verify the utility of *COI* genes as probes and the possibility to identify all the Korean freshwater fish species that will be exploited in future studies. If various DNA chips are arrayed using different species-specific oligonucleotide probes designed for a specific application, these chips will be more valuable techniques for biodiversity studies and environmental monitoring.

Materials and Methods

Target and non-target species

Thirty-one species and their 325 *COI* barcodes were chosen from the *COI* barcode dataset of 53 Korean freshwater fish species previously published[18]. These sequences were aligned using MUSCLE software and exposed to Nexus format[26]. Sequence divergences among individuals were quantified using the Kimura-2 parameter (K2P) distance model[27], and a neighbor-joining (NJ) tree of K2P distance was created to provide a graphic representation of divergence patterns among species using MEGA version 4.0.2[28]. Of these 31 species, 11 target species were used to construct oligonucleotide probes and arrayed in a DNA chip. Selected target species were completed by assessing morphological indistinctness in either the mature or immature state of *R. notatus, R. uyekii, A. gracilis, A. rhombeus, A. yamatsutae, O. uncirostris amurensis, Z. platypus, Z. koreanus, C. auratus, R. brunneus*, and *R. giurinus*. The 20 species were designated as non-target species and DNAs were used as confirmatory non-bindings to target oligonucleotide probes. These species included *Eryghroculter erythropterus,*

Table 2. Summary of 31 Korean freshwater fish species dataset.

Species	No. of sequence	Accession numbers of GenBank	Target (T)/non-Target (N) species
Abbottina rivularis	13	HQ536228-HQ536240	N
Acanthogobius hasta	7	HQ536241-HQ536247	N
Acanthorhodeus gracilis	4	HQ536248-HQ536251	T
Acheilognathus rhombeus	13	HQ536252-HQ536264	T
Acheilognathus yamatsutae	15	HQ536267-HQ536281	T
Carassius auratus	26	HQ536290-HQ536315	T
Chaenogobius urotaenia	4	HQ536320-HQ536323	N
Coreoleuciscus splendidus	5	HQ536327-HQ536331	N
Erythroculter erythropterus	5	HQ536348-HQ536352	N
Gnathopogon strigatus	15	HQ536353-HQ536367	N
Hemibarbus labeo	10	HQ536370-HQ536379	N
Liobagrus andersoni	3	HQ536394-HQ536396	N
Microphysogobio yaluensis	5	HQ536399-HQ536403	N
Opsariichthys uncirostris amurensis	11	HQ536411-HQ536421	N
Oryzias sinensis	7	HQ536422-HQ536428	T
Phoxinus oxycephalus	5	HQ536429-HQ536433	N
Phoxinus phoxinus	3	HQ536434-HQ536436	N
Pseudorasbora parva	15	HQ536439-HQ536453	N
Rhinogobius brunneus	15	HQ536455-HQ536469	T
Rhinogobius giurinus	11	HQ536470-HQ536480	T
Rhodeus notatus	19	HQ536481-HQ536499	T
Rhodeus uyekii	3	HQ536505-HQ536507	T
Rhynchocypris kumgangensis	2	HQ536508, HQ536509	N
Rhynchocypris steindachneri	2	HQ536404, HQ536405	N
Silurus microdorsalis	1	HQ536514	N
Takifugu obscurus	1	HQ536515	N
Trachidermus fasciatus	4	HQ536516-HQ536519	N
Tridentiger brevispinis	7	HQ536525-HQ536531	N
Tridentiger obscurus	6	HQ536532-HQ536537	N
Zacco koreanus	63	HQ536538-HQ536600	T
Zacco platypus	25	HQ536601-HQ536625	T

Tridentiger obscures, Rhynchocypris kumgangensis, Trachidermus fasciatus, Chaenogobius urotaenia, Hemibarbus labeo, Oryzias sinensis, Microphysogobio yaluensis, Silurus mierodorsalis, Tridentiger brevispinis, Rhynchocypris steindachneri, Abbottina rivularis, Phoxinus oxycephalus, Coreoleuciscus splendidus, Phoxinus phoxinus, Gnathopogon strigatus, Pseudorasbora parva, Liobagrus andersoni, Acanthogobius hasta, and *Takifugu obscures* (Table 2).

DNA chip fabrication

DNA chip was fabricated (Figure 1A) using target species-specific probes mixed with 2× spotting buffer containing 6× standard saline citrate (SSC) and 3M betaine, which was printed and synthesized onto 5 chamber CSS-100 silylated slides (Cell Associate, USA) at 25°C and 60% humidity. The DNA chip was washed with 0.1% sodium dodecyl sulfate (SDS) for 5 min and incubated in 375 mL of sodium borohydride solution ($NaBH_4$, 1.3 g; phosphate-buffered saline [PBS], 125 mL; and 100% ethanol) for 5 min. Finally, the slide was washed twice for 5 min using sterile double-distilled water and centrifuged at 800 rpm for 5 min.

DNA chip experiment

There are 6 main steps for conducting DNA chip experiments, which include preparation of a hybridization solution mixed with PCR product and hybridization, incubation, washing, scanning, and image analysis as described in Figure 1B. In the hybridization step, Cy3-labeled amplified PCR products from test species were hybridized in 100 μL of 3× SSC, 0.3% SDS for 1 h incubation at 55°C. After hybridization, DNA chips were immediately washed in a 1× SSC containing 0.1% SDS for 5 min, then washed twice using 0.1× SSC for 3 min and centrifuged at 800 rpm for 5 min. Next, hybridization signals were detected using a GenePix 4000B (Axon instrument, USA) at a photomultiplier tube (PMT) gain of 600 and 99% laser power. After scanning, signal intensities of each spot were calculated by GenePix Pro 4.1 software.

Acknowledgements This work was supported by a grant from the Eco-Technopia 21 Project funded by the Ministry of Environment of Korea (2008-05002-0065-0), the second stage of the Brain Korea 21 Project in 2009, and 2008-2009 Sangji University Sabbatical Leave Grant.

References

1. Tautz, D., Arctander, P., Minelli, A., Thomas, R. & Vogler, A. DNA points the way ahead in taxonomy. *Nature* **418**, 479-479 (2002).
2. Hebert, P. & Gregory, T. The promise of DNA barcoding for taxonomy. *Syst. Biol.* **54**, 852-859 (2005).
3. Schindel, D. & Miller, S. DNA barcoding a useful tool for taxonomists. *Nature* **435**, 17 (2005).
4. Hajibabaei, M., Janzen, D., Burns, J., Hallwachs, W. & Hebert, P. DNA barcodes distinguish species of tropical Lepidoptera. *PNAS* **103**, 968-971 (2006).
5. Rock, J. *et al*. DNA barcodes of fish of the Scotia Sea, Antarctica indicate priority groups for taxonomic and systematics focus. *Antarctic Sci.* **20**, 253-262 (2008).
6. Stahls, G. *et al*. *COI* barcodes for identification of Merodon hoverflies (Diptera, Syrphidae) of Lesvos Island, Greece. *Mol. Ecol. Resour.* **9**, 1431-1438 (2009).
7. Hebert, P., Ratnasingham, S. & de Waard, J. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. B* **270**, 96-99 (2003).
8. Hubert, N. *et al*. Identifying Canadian freshwater fishes through DNA barcodes. *PloS One* **3**, e2490 (2008).
9. Moritz, C. & Cicero, C. DNA barcoding: promise and pitfalls. *PLoS Biol.* **2**, 1529-1531 (2004).
10. Ward, R., Hanner, R. & Hebert, P. The campaign to DNA barcode all fishes, FISH-BOL. *J. Fish Biol.* **74**, 329-356 (2009).
11. Ratnasingham, S. & Hebert, P. BOLD: the Barcode of Life Data System (http://www. barcodinglife. org). *Mol. Ecol. Resour.* **7**, 355-364 (2007).
12. Song, K., Jung, J., Koo, H. & Kim, W. Development of species-specific molecular marker as a tool for discrimination between Crucian Carp Gengorobuna (Carassius cuvieri) introduced from Japan and Korean native one (C. auratus). *Korean J. Limnol.* **40**, 143-148 (2007).
13. Kochzius, M. *et al*. DNA microarrays for identifying fishes. *Mar. Biotechnol.* **10**, 207-217 (2008).
14. Altschul, S., Gish, W., Miller, W., Myers, E. & Lipman, D. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410 (1990).
15. Steinke, D., Vences, M., Salzburger, W. & Meyer, A. TaxI: a software tool for DNA barcoding using distance methods. *Phil. Trans. R. Soc. B* **360**, 1975-1980 (2005).
16. DasGupta, B., Konwar, K., Mandoiu, I. & Shvartsman, A. DNA-BAR: distinguisher selection for DNA barcoding. *Bioinformatics* **21**, 3424-3426 (2005).
17. Rach, J., DeSalle, R., Sarkar, I., Schierwater, B. & Hadrys, H. Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. *Proc. Biol. Sci.* **275**, 237-247 (2008).
18. Kim, S., Eo, H., Koo, H., Choi, J. & Kim, W. DNA barcode-based molecular identification system for fish species. *Mol. Cells* **30**, 1-10 (2010).
19. Pfunder, M., Holzgang, O. & Frey, J. Development of microarray-based diagnostics of voles and shrews for use in biodiversity monitoring studies, and evaluation of mitochondrial cytochrome oxidase I vs. cytochrome b as genetic markers. *Mol. Ecol.* **13**, 1277-1286 (2004).
20. Engelmann, J. *et al*. Modelling cross-hybridization on phylogenetic DNA microarrays increases the detection power of closely related species. *Mol. Ecol. Resour.* **9**, 83-93 (2009).
21. Chung, I. *et al*. A DNA microarray for identification of selected Korean birds based on mitochondrial cytochrome c oxidase I gene sequences. *Mol. Cells* **30**, 295-301 (2010).
22. Park, J. *et al*. A DNA microarray for species identification of cetacean animals in Korean water. *BioChip J.* **4**, 197-203 (2010).
23. Summerbell, R. *et al*. Microcoding: the second step in DNA barcoding. *Phil. Trans. R. Soc. B* **360**, 1897-1903 (2005).
24. Hajibabaei, M., Singer, G., Clare, E. & Hebert, P. Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. *BMC Biology*. **5**, 24 (2007).
25. Lee, H. *et al*. G. DNA chip evaluation as a diagnostic device. *Current Applied Physics* **5**, 433-437 (2005).
26. Edgar, R. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-1797 (2004).
27. Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111-120 (1980).
28. Tamura, K., Dudley, J., Nei, M. & Kumar, S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596-1599 (2007).