SCIENCE CHINA
Life Sciences

• RESEARCH PAPER •

January 2012 Vol.55 No.1: 80–88
doi: 10.1007/s11427-012-4266-2

Selection of a DNA barcode for Nectriaceae from fungal whole-genomes

ZENG ZhaoQing[1,2], ZHAO Peng[1], LUO Jing[1], ZHUANG WenYing[1*] & YU ZhiHe[3]

[1]*State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China;*
[2]*Graduate University of Chinese Academy of Sciences, Beijing 100049, China;*
[3]*College of Life Sciences, Yangtze University, Jingzhou 434025, China*

Received October 21, 2011; accepted November 17, 2011

A DNA barcode is a short segment of sequence that is able to distinguish species. A barcode must ideally contain enough variation to distinguish every individual species and be easily obtained. Fungi of Nectriaceae are economically important and show high species diversity. To establish a standard DNA barcode for this group of fungi, the genomes of *Neurospora crassa* and 30 other filamentous fungi were compared. The expect value was treated as a criterion to recognize homologous sequences. Four candidate markers, *Hsp90*, *AAC*, *CDC48*, and *EF3*, were tested for their feasibility as barcodes in the identification of 34 well-established species belonging to 13 genera of Nectriaceae. Two hundred and fifteen sequences were analyzed. Intra- and inter-specific variations and the success rate of PCR amplification and sequencing were considered as important criteria for estimation of the candidate markers. Ultimately, the partial *EF3* gene met the requirements for a good DNA barcode: No overlap was found between the intra- and inter-specific pairwise distances. The smallest inter-specific distance of *EF3* gene was 3.19%, while the largest intra-specific distance was 1.79%. In addition, there was a high success rate in PCR and sequencing for this gene (96.3%). *CDC48* showed sufficiently high sequence variation among species, but the PCR and sequencing success rate was 84% using a single pair of primers. Although the *Hsp90* and *AAC* genes had higher PCR and sequencing success rates (96.3% and 97.5%, respectively), overlapping occurred between the intra- and inter-specific variations, which could lead to misidentification. Therefore, we propose the *EF3* gene as a possible DNA barcode for the nectriaceous fungi.

barcoding gap, expect value, fungal genomes, homologous sequence, PCR and sequencing success rate, sequence variation

Citation: Zeng Z Q, Zhao P, Luo J, *et al*. Selection of a DNA barcode for Nectriaceae from fungal whole-genomes. Sci China Life Sci, 2012, 55: 80−88, doi: 10.1007/s11427-012-4266-2

DNA barcoding, in the strict sense, is defined as the standardized analysis of an easily obtained fragment for sequence-based identification of species [1]. A barcode should have the attributes required for rapid species identification for biodiversity assessment, detection of species invasion, food and feedstuff safety, ecology, natural resources conservation, ornamental fish trade, and human health [2–12]. In the animal kingdom, the mitochondrial cytochrome c oxidase 1 gene is capable of species identification [13–18]. In plants, the combination of *rbcL* and *matK* [19], *psbA-trnH* and ITS (internal transcribed spacer) [20] and ITS2 [21,22] are considered as appropriate barcodes. However, recognition of a universal DNA barcode for fungi has been a more challenging task, because many fungi are microscopic, species concepts are inconsistently applied, and some fungi are pleomorphic [23]. In recent years, studies have increasingly focused on fungal DNA barcoding. The ITS region was found to be effective for species discrimination of six genera in Zygomycota [24], *Trichoderma* and *Hypocrea* in Ascomycota [25], and the *Cortinarius* section

*Corresponding author (email: zhuangwy@im.ac.cn)

of *Calochroi* [26], *Melampsora* [27] and some edible fungi [23] in Basidiomycota. Nevertheless, this region appeared to be problematic in dealing with some closely related species of arbuscular mycorrhizal fungi [1], the blue stain fungi [28], and nectriaceous fungi [29]. Other barcode loci, such as nuclear 28S rDNA and translation elongation factor 1α gene, provide a greater capability of species separation in some groups [25,30,31]. Zhao *et al*. [29] investigated four candidate DNA barcode markers for their feasibility in the identification of some nectriaceous fungi, and the *β-tubulin* gene was proposed as a possible barcode. So far, an official DNA barcode for fungi has not been defined [1,32]. In previous studies, the majority of the candidate markers were selected from genes applied to fungal phylogeny.

Since the publication of the first fungal genome, the model yeast *Saccharomyces cerevisiae* in 1996 [33], more than 400 fungal genome projects have been launched. One hundred and twenty-eight fungal species have been sequenced or partially sequenced [34]. Recent research on fungal genomics have concentrated on phylogeny [35,36], pathogenicity [37,38], antifungal drug discovery [39–41], yeast comparative genomics [42,43], and the search for a DNA barcode [44,45]. Comparative genomics has become a powerful and useful tool as datasets increase, and more data are able to discriminate conserved from divergent DNA [46]. Moreover, this approach identified conserved genes that are only found in fungi [42].

Fungi of Nectriaceae are distributed worldwide and occur on various substrates in tropical and subtropical regions [47]. They are economically important, show very high species diversity, and a broad range of lifestyles [48]. Rapid species identification of the group is required for plant disease diagnoses, discovery of new bioactive compounds, exploration of potential biocontrol agents, and protection from harmful mycotoxins [29]. In this study, we compared the genomes of *Neurospora crassa* and 30 other filamentous fungi, and utilized the expect value (*E*-value) as a criterion to identify homologous sequences. Among the obtained sequences, four candidate barcode markers were selected and tested for their feasibility as a DNA barcode for the nectriaceous fungi. The genes encode the heat shock protein 90 (*Hsp90*), the ADP/ATP carrier protein (*AAC*), the cell division cycle protein 48 (*CDC48*), and the elongation factor 3 (*EF3*).

1 Materials and methods

1.1 Materials studied

Two hundred and fifteen sequences of four candidate markers, *Hsp90*, *AAC*, *CDC48*, and *EF3* genes, from 34 clearly documented and generally accepted species belonging to 13 genera of Nectriaceae were analyzed. All sequences have been deposited in GenBank (Table 1).

1.2 Genome comparison

To select a candidate DNA barcode marker, 31 genomes of filamentous fungi (including 16 spp. of *Eurotiomycetes*, 10 of *Sordariomycetes*, three of *Dothideomycetes*, and two of *Leotiomycetes*), which are publicly available on the Internet, were investigated. The program Standalone BLAST 2.2.21 [49] was set up on a Windows XP system, and a local database was generated for each genome of the filamentous fungi. The complete genome of *Neurospora crassa* was split into individual gene sets. Each protein-coding gene was compared separately against each of the 30 genomes using Standalone BLASTN. An *E*-value cutoff of 0.1 was used to identify significant hits in this analysis. All BLAST analyses were run with default parameters in Standalone BLAST. The BLAST output files were parsed using PERL scripts, which generated a single line of output from each BLAST output file [42,50]. For each dataset, the individual lines were combined into a single file, and the file imported into a spreadsheet program, and then analyzed using Microsoft Office Excel (Windows XP). Certain specific groups of homologous sequences with a given *E*-value were aligned by ClustalX 1.81 [51].

1.3 Primer design

To design primers for PCR amplification and sequencing of the candidate markers, we used the corresponding sequences of gene fragments derived from whole-genome sequences of certain Sordariomycetes species including *Fusarium graminearum*, *F. oxysporum*, *F. verticillioides*, *Magnaporthe grisea*, *N. crassa*, *Verticillium albo-atrum*, and *V. dahliae*. A complete alignment was carried out using ClustalX [51] to find the conserved regions for primer design. Five primer pairs were designed using Primer Premier 5.0 [52], according to the main principles proposed by Compton [53], Dieffenbach and Dveksler [54], Goller *et al*. [55], Innis *et al*. [56], Saiki [57], and Sambrook and Russell [58].

1.4 DNA amplification and sequencing

Genomic DNA of each strain was isolated from mycelium grown on potato dextrose agar at room temperature for about two weeks [59]. The PCR reaction was performed on an ABI 2720 Thermal Cycler (Gene Co. Ltd., Foster City, California, USA). The 25 μL reaction system comprised 16.25 μL of double distilled water, 2.5 μL of 10× PCR buffer, 2 μL of $MgCl_2$ (25 mmol L^{-1}), 1 μL of each primer (10 μmol L^{-1}), 0.5 μL of dNTP (10 mmol L^{-1} each), 1.5 μL of DNA template, and 0.25 μL Taq DNA polymerase (5 U μL^{-1}). The PCR primers used were newly designed in this study. The cycling conditions were an initial step of 5 min at 95°C; 35 cycles of 40 s at 94°C, 40 s at 54/56/52/62°C for *Hsp90*/*AAC*/*CDC48*/*EF3*, respectively, and 40 s at 72°C; followed by 10 min at 72°C. The PCR products were verified by electrophoresis of 2.5 μL products on a 1% agarose gel and staining with ethidium bromide. A molecular size marker was included to estimate length of the amplification

Table 1 Materials used in this study

Species	Collection number or source[a)]	GenBank accession number			
		Hsp90	*AAC*	*CDC48*	*EF3*
Albonectria rigidiuscula (Berk. & Broome) Rossman & Samuels	HMAS 183135	JN131641	JN131694	JN131748	JN131802
	HMAS 183509	JN131642	JN131695	JN131749	JN131803
Chaetopsinectria chaetopsinae (Samuels) J. Luo & W.Y. Zhuang	HMAS 76860	JN131658	JN131712	JN131766	JN131820
Corallomycetella repens (Berk. & Broome) Rossman & Samuels	AR 4659	JN131615	JN131668	JN131722	JN131776
Cosmospora coccinea Rabenh.	CBS 114050	JN131609	JN131662	JN131716	JN131770
C. cupularis J. Luo & W.Y. Zhuang	HMAS 97514	JN131638	JN131691	JN131745	JN131799
C. episphaeria (Tode) Rossman & Samuels	HMAS 99194	JN131632	JN131685	JN131739	JN131793
C. henanensis Y. Nong & W.Y. Zhuang	HMAS 183528	JN131635	JN131688	JN131742	JN131796
C. meliopsicola (Henn.) Rossman & Samuels	HMAS 86473	JN131618	JN131671	JN131725	JN131779
Cosmospora sp. 1	HMAS 76861	JN131655	JN131709	JN131763	JN131817
	CBS 122578	JN131610	JN131663	JN131717	JN131771
Cosmospora sp. 2	GJS 85.205	JN131649	JN131703	JN131757	JN131811
C. vilior (Starbäck) Rossman & Samuels	HMAS 183636	JN131646	JN131700	JN131754	JN131808
Cyanonectria cyanostoma (Sacc. & Flageolet) Samuels & Chaverri	GJS 98.127	JN131652	JN131706	JN131760	JN131814
Gibberella zeae (Schwein.) Petch	HMAS 98297	JN131631	JN131684	JN131738	JN131792
Haematonectria haematococca (Berk. & Broome) Samuels & Rossman	HMAS 183156	JN131636	JN131689	JN131743	JN131797
	HMAS 183157	JN131637	JN131690	JN131744	JN131798
	HMAS 183514	JN131643	JN131696	JN131750	JN131804
H. ipomoeae (Halst.) Samuels & Nirenberg	HMAS 188477	JN131647	JN131701	JN131755	JN131809
	HMAS 188475	JN131648	JN131702	JN131756	JN131810
Lanatonectria flavolanata (Berk. & Broome) Samuels & Rossman	HMAS 91516	JN131622	JN131675	JN131729	JN131783
L. flocculenta (Henn. & E. Nyman) Samuels & Rossman	HMAS 76873	JN131653	JN131707	JN131761	JN131815
Nectria australiensis Seifert	HMAS 83397	JN131654	JN131708	JN131762	JN131816
N. cinnabarina (Tode) Fr.	HMAS 98306	JN131625	JN131678	JN131732	JN131786
	HMAS 98311	JN131629	JN131682	JN131736	JN131790
N. pseudotrichia (Schwein.) Berk. & M.A. Curtis	HMAS 97518	JN131634	JN131687	JN131741	JN131795
	HMAS 183560		JN131699	JN131753	JN131807
	HMAS 183175	JN131645	JN131698	JN131752	JN131806
	HMAS 183559	JN131644	JN131697	JN131751	JN131805
Neonectria castaneicola (W. Yamam. & Oyasu) Tak. Kobay. & Hirooka	HMAS 76865	JN131656	JN131710	JN131764	JN131818
	HMAS 83369	JN131657	JN131711	JN131765	JN131819
	HMAS 183542	JN131640	JN131693	JN131747	JN131801
N. confusa J. Luo & W.Y. Zhuang	HMAS 99197	JN131624	JN131677	JN131731	JN131785
	HMAS 99198	JN131626	JN131679	JN131733	JN131787
N. discophora var. *discophora* (Mont.) Mantiri & Samuels	HMAS 98333	JN131619	JN131672	JN131726	JN131780
	HMAS 98327	JN131621	JN131674	JN131728	JN131782
N. ditissimopsis P. Zhao, J. Luo & W.Y. Zhuang	HMAS 98328	JN131627	JN131680	JN131734	JN131788
	HMAS 99206	JN131628	JN131681	JN131735	JN131789
	HMAS 98329	JN131630	JN131683	JN131737	JN131791
N. hubeinensis W.Y. Zhuang, Y. Nong & J. Luo	HMAS 98331	JN131620	JN131673	JN131727	JN131781
N. jungneri Henn.	GJS 08-233	JN131607	JN131660	JN131714	JN131768
N. major (Wollenw.) Castl. & Rossman	HMAS 183184	JN131639	JN131692	JN131746	JN131800
N. ramulariae Wollenw.	CBS 182.36	JN131612	JN131665	JN131719	JN131773
	CBS 151.29	JN131611	JN131664	JN131718	JN131772
N. shennongjiana J. Luo & W.Y.Zhuang	HMAS 183185	JN131659	JN131713	JN131767	JN131821
N. sinensis J. Luo & W.Y. Zhuang	HMAS 183186	JN131651	JN131705	JN131759	JN131813
N. veuillotiana (Sacc. & Roum.) Mantiri & Samuels	HMAS 98332	JN131623	JN131676	JN131730	JN131784
	HMAS 99207	JN131633	JN131686	JN131740	JN131794
	GJS 91-116	JN131650	JN131704	JN131758	JN131812
Pseudonectria pachysandricola B.O. Dodge	CBS 476.92	JN131616	JN131669	JN131723	JN131777
	CBS 501.63	JN131617	JN131670	JN131724	JN131778
P. rousseliana (Mont.) Wollenw.	AR 2714	JN131613	JN131666	JN131720	JN131774
	CBS 114049	JN131608	JN131661	JN131715	JN131769
Rubrinectria olivacea (Seaver) Rossman & Samuels	AR 4331	JN131614	JN131667	JN131721	JN131775

a) CBS, Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands; GJS, G.J. Samuels; HMAS, Herbarium of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China; AR, Amy Rossman.

products. The obtained amplicons were purified by PCR Product Purification Kit (Biocolor BioScience & Technology Company) and sequenced in both directions, using the same primers used for PCR amplification, on an ABI 3730 XL DNA Sequencer (SinoGenoMax Co., Ltd.). The sequencing reactions were performed according to the manufacturer's protocol.

The original forward and reverse sequences were assembled manually, aligned using ClustalX [51], and edited to adjust the aligned sequences by BioEdit 7.0 [60]. The non-alignable sequences for *Hsp90*, *AAC*, and *CDC48* gene fragments were removed to decrease the influence of large gaps. No gaps were found in the *EF3* gene.

1.5 Comparison of intra- and inter-specific divergences

The aligned sequences were input into DNAstar 7.1.0 (Lasergene, WI, USA) to calculate the similarity matrices. A visualization analysis tool, TaxonGap 2.4.1 [61], was used to illustrate the sequence divergences within and between species of the candidate markers for each of the 34 investigated species.

The intra- and inter-specific pairwise distances were calculated using the K2P model in MEGA 4.1 [62]. The frequency distributions of the intra- and inter-specific distances were analyzed to check the barcoding gap, i.e., the space between intra- and inter-specific distances [63], using Microsoft Office Excel.

1.6 Success rates of test barcode sequence acquisition

The success rates of PCR amplification and sequencing were evaluated. PCR amplification was considered as successful when there was a single and clear band of the expected size on agarose gels. A high quality chromatogram counted as successful sequencing. The success rate of PCR amplification multiplied by that of sequencing determined the success rate of PCR amplification and sequencing.

2 Results

Thirty-one genomes of filamentous fungi were compared. Homologous sequences with *E*-values greater than 10^{-30} showed a high proportion of polymorphism, based on alignments using ClustalX [51]. Of the homologous sequences with *E*-values near or less than 10^{-100}, the *Hsp90*, *AAC*, *CDC48*, and *EF3* genes were randomly selected as candidate markers to investigate their feasibility in the identification of the 34 nectriaceous species. This is the first time that these gene fragments have been selected as candidate markers for the purpose of fungal DNA barcoding.

The designed primers are shown in Table 2. Degenerate primers were devised for the *Hsp90*, *EF3*, and *CDC48* genes because variable bases were found at some positions. PCR amplifications and sequencing were carried out successfully for all the selected gene fragments. The sequence lengths were 557–589 base pairs (bp) for *Hsp90* gene, 551–607 bp for *AAC* gene, 420–454 bp for *CDC48* gene, and 501 bp for *EF3* gene.

To judge the suitability of a DNA barcode marker, the intra- and inter-specific sequence divergence is a very important criterion. Comparisons of the four candidate markers for each of the 34 Nectriaceae species were performed by TaxonGap [61]. The results are summarized in Figure 1. For each species, sequence similarity of the same gene within the same species was high; therefore, the relevant intra-specific variation (shown as grey bars in Figure 1) was low. For the *CDC48* and *EF3* genes, the inter-specific variations were very similar, and are apparently higher than those for the *Hsp90* and *AAC* genes, and there was a clear gap between the maximum intra-specific variations and the minimum inter-specific variations. In the case of the *Hsp90* gene, the intra-specific variation for *Nectria cinnabarina* was very close to the inter-specific variation between *Haematonectria haematococca* and *H. ipomoeae*. Similarly, the intra-specific variation for *Albonectria rigidiuscula* was almost the same as the inter-specific variation between *Cosmospora henanensis* and *Neonectria confusa* when the *AAC* gene was used, which indicates this gene may cause incorrect identification.

The frequency distributions of the intra- and inter-specific pairwise distances of the four candidates were analyzed to check the barcoding gap [63]. The results are shown in Figure 2. In the cases of the *CDC48* and *EF3* genes, the genetic distances within any single species did not exceed those among species, and a distinct gap was present between the inter- and intra-specific distances. In contrast, overlapping occurred when using the *Hsp90* and *AAC* genes, which indicated that they are not useful as barcode markers.

The ease of PCR amplification and nucleotide sequence acquisition is also important in evaluating a DNA barcode marker. Three of the four candidate markers were readily amplified and sequenced, with high success rates ($\geqslant$96.3%) (Table 3). However, for the *CDC48* gene, additional primers were required.

Table 2 Primers used in this study

Primer name	Primer sequence (5′→3′)
Hsp90F	CATCATCAACACHGTCTACTCC
Hsp90R	KGAGATRAACTCGGAGTGCTTC
AACF	CCAACGTCATCCGTTACTT
AACR	ACCCTTGAAGAGAGACTTGA
CDCF	GCCGTCAACGATGARAAC
CDCR	CGACKACYTTGAACTCTACT
CDC1F	CGACTGCGATCCTCAAG
CDC1R	CCCTCGCAGTGAATGAC
EF3F	GACCACCATTGACTGGACCA
EF3R	TTGGAGGTRCCAGGGTACT

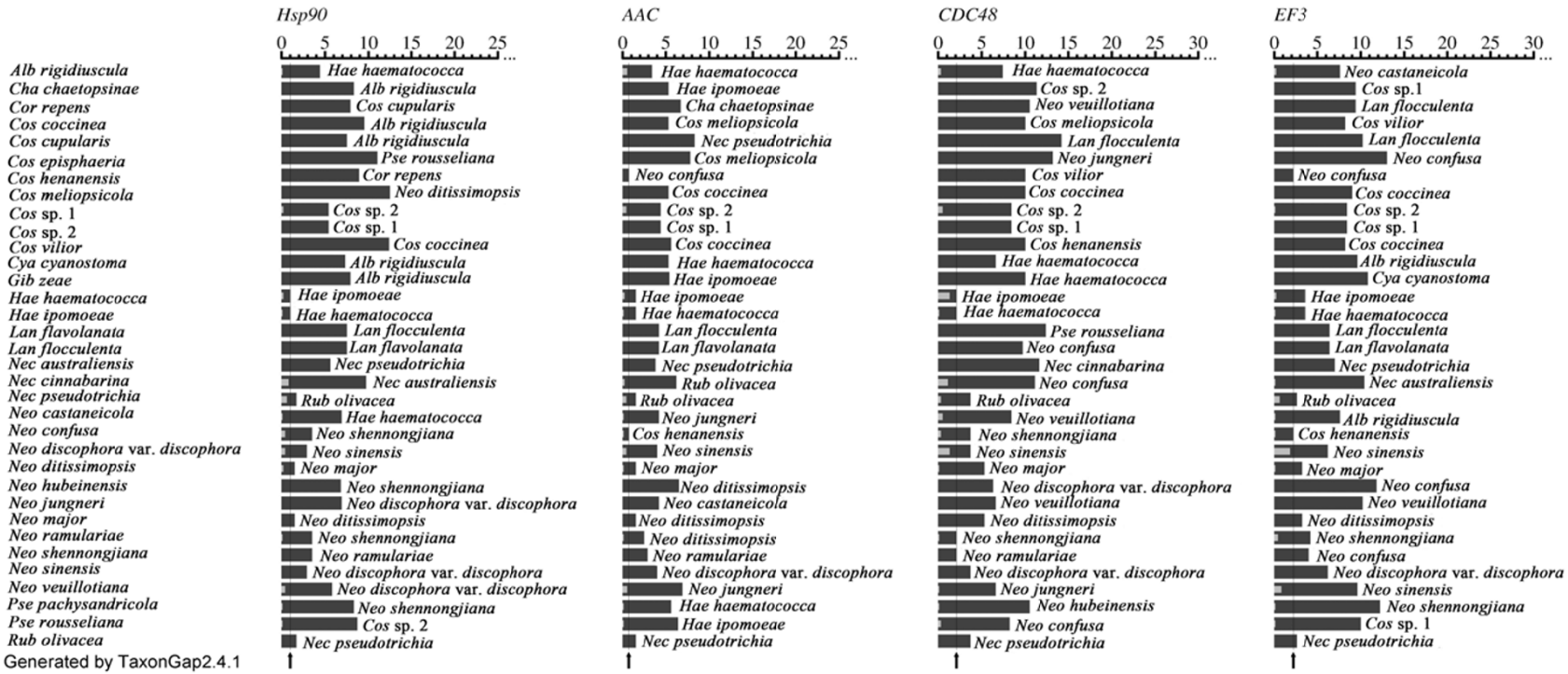

Figure 1 Comparisons of intra- and inter-specific variations among *Hsp90*, *AAC*, *CDC48*, and *EF3* genes of the nectriaceous fungi generated by the software TaxonGap. The grey and black bars represent the intra- and inter-specific variations, respectively. The thin, black lines indicate the smallest inter-specific variation. Names next to the dark bars indicate the closest species to that listed on the left.

Table 3 Success rates of PCR and sequencing of *Hsp90*, *AAC*, *CDC48*, and *EF3* genes of 34 species of Nectriaceae

Candidate barcode	*Hsp90*	*AAC*	*CDC48*	*EF3*
Efficiency of PCR amplification	100%	97.5% (79/81)	87.7% (71/81)	96.3% (78/81)
Success rates of sequencing	96.3% (78/81)	100%	95.8% (68/71)	100%
Success rate of PCR and sequencing	96.3%	97.5%	84.0%	96.3%

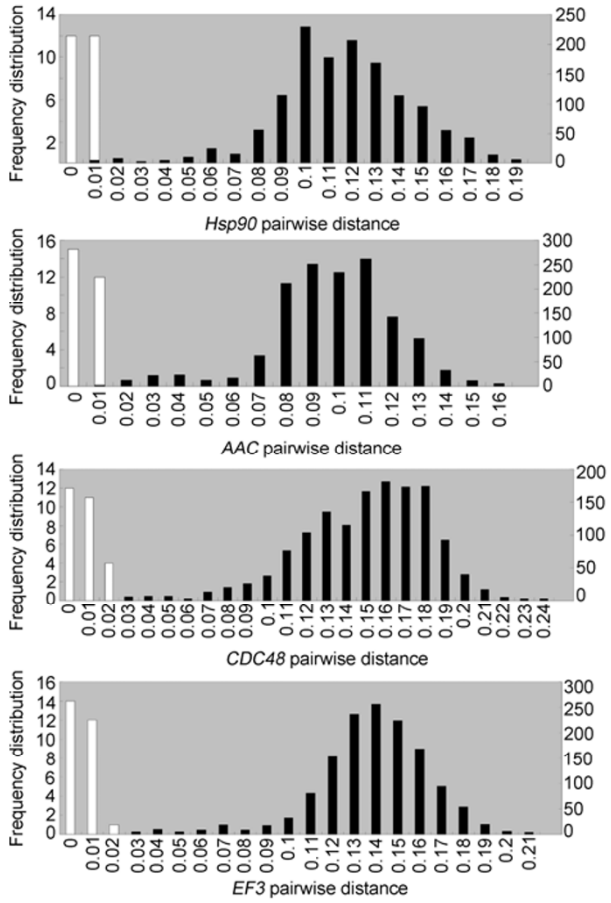

Figure 2 Comparisons of frequency distribution of intra- and inter-specific pairwise distances among *Hsp90*, *AAC*, *CDC48*, and *EF3* genes for the nectriaceous fungi generated with MEGA and Excel. The intra- and inter-specific distances are shown as white and black bars, respectively.

3 Discussion

3.1 Selection of genes for DNA barcoding

E-value is a parameter that describes the number of hits one can "expect" to see by chance when searching a database of a particular size [64]. Following the study by Fitzpatrick *et al*. [36], the top BLAST hits longer than 300 nucleotides were retained as putative open reading frames for further analysis. There has been some discussion concerning the statistical significance represented by *E*-value. Pertsemlidis and Fondon [65] and Zeng *et al*. [66] treated *E*-values less than 0.1 as a measure of statistical significance, while Pearson [67] suggested 0.02 as the boundary measurement. *E*-values below 10^{-5} are often considered to represent homology [42,68–71]. In the present analysis, an *E*-value cutoff of 0.1 was first chosen to identify significant hits. Our results indicated that a threshold of 10^{-5} is better than 0.1 for searching for homologs among the fungi studied. Variability of the obtained homologous sequences was analyzed, based on alignment using ClustalX [51], to find the candidate markers. The result indicated that relatively conserved homologous sequences with *E*-values around or less than 10^{-100} may be suitable as candidate markers. As a result, four genes were selected as candidate markers.

Non-coding regions, such as the ITS, often have insufficient phylogenetic information to unequivocally identify closely related species in some genera of Ascomycota [29,31,72,73]. In contrast, protein-coding genes have higher information contents [74]. Accordingly, protein-coding DNA sequences might represent better candidates for fungal DNA barcoding.

3.2 Evaluation of the candidate DNA markers

Many criteria have been used to determine an ideal DNA barcode, such as a short fragment, universally used, having adequate variations among species, and conserved within a species, exhibiting a high degree of species resolution [19,75]. Sequence variations, frequency distribution of intra- and inter-specific pairwise distances, and easiness of sequence acquisition were treated as essential.

The *Hsp90*, *AAC*, *CDC48*, and *EF3* genes were tested for their feasibility as a DNA barcode for the nectriaceous fungi. The single copy gene *EF3* encoding the elongation factor 3 is a unique, essential, and soluble component of the translational system in fungi [76–78]. Our study presents a strong case for the partial *EF3* gene being the most promising DNA barcode for this group. First, it exhibits the smallest sequence divergence within an individual species and a maximal separation among species. As calculated by TaxonGap [61], all the inter-specific variations were greater than the intra-specific variations, and a clear gap existed between them (Figure 1). This was further substantiated by the frequency distribution of the intra- and inter-specific pairwise distances calculated using MEGA [62] and Microsoft Office Excel. Its smallest inter-specific distance was 3.19%, and the largest intra-specific distance was 1.79%; thus, *EF3* appeared to possess the appropriate intra- and inter-specific variations. No overlapping occurred between the intra- and inter-specific pairwise distances (Figure 2). Secondly, the partial *EF3* gene lacks an intron and is relatively easy to obtain. Its PCR amplification and sequencing

success rate reached 96.3%, which was slightly lower than that of the *AAC* gene (97.5%) (Table 3).

The *CDC48* gene exists as a single copy per haploid genome [79]. Its encoded protein plays an essential role in cell proliferation, cell cycle progression, and ATP-dependent fusion of endoplasmic reticular membranes [80,81]. Determination of genetic divergence using two different methods confirmed that the *CDC48* region possesses a good sequence divergence within and among species, which is equally high as that shown by the *EF3* gene. An obvious boundary occurred between the intra- and inter-specific pairwise distances (Figure 2). However, it had a lower PCR and sequencing success rate (84%, Table 3) when using a single pair of primers. Additional primers (CDC1F and CDC1R) were required to reach a higher success rate (Table 2). The Hsp90 protein plays a key role in signal transduction, cell cycle control, protein folding, protein degradation, cell signaling, and morphological evolution [82]. The nuclear gene *Hsp90* is a single-copy gene and usually contains multiple introns [83,84]. It is highly conserved and extensively used in phylogenetic analyses [85–87]. In this study, although it had a high PCR and sequencing success rate (96.3%) (Table 3), it cannot separate the intra- and inter-specific variations adequately (Figure 1), which influences its function as a barcode.

AAC occurs as a single copy in the genome [88]. The protein is located on the inner membrane of the mitochondria and catalyses the exchange diffusion of ADP and ATP [89]. Our study showed that this gene had the highest PCR and sequencing success rate (97.5%) (Table 3). However, it showed a relatively low inter-specific variation leading to inaccurate identification of species (Figure 1).

Our results indicate that the partial *EF3* gene meets the requirements for a good DNA barcode. *CDC48* gene possessed good sequence variations among species, but the PCR and sequencing success rate was relatively low. The *Hsp90* and *AAC* genes had high PCR and sequencing success rates, while overlapping occurred between the intra- and inter-specific distances, which may lead to misidentification. Compared with *β-tubulin*, which was previously suggested as a possible DNA barcode for the nectriaceous fungi [29], the sequence acquisition of *EF3* gene is easier.

Recently, Robert *et al*. [44] devised an approach to locate potential barcode markers from fungal genomes. Lewis *et al*. [45] managed to identify a barcode gene using a taxonomy-aware processing pipeline. Their results indicated that genome-mining has a potential use in fungal DNA barcoding. Our study may represent the first step towards selecting DNA barcodes from fungal whole-genome comparisons. With the increasing availability of genome datasets, we believe that comparative genomics will play an essential role in DNA barcoding.

This work was supported by the National Natural Science Foundation of China (Grant No. 31070015), the Special Project for Fundamental Research (Grant No. 2006FY120100) from Ministry of Science and Technology of China, and the Knowledge Innovation Program of Chinese Academy of Sciences (Grant No. KSCX2-EW-J-6). We are very grateful to Prof. Tom Hsiang for his review of the manuscript and language corrections, and to Dr. Zheng HuanDi for technical help.

1 Stockinger H, Krüger M, Schüßler A. DNA barcoding of arbuscular mycorrhizal fungi. New Phytol, 2010, 187: 461–474
2 Armstrong K F, Ball S L. DNA barcodes for biosecurity: invasive species identification. Phil Trans R Soc B, 2005, 360: 1813–1823
3 Ball S L, Armstrong K F. DNA barcodes for insect pest identification: a test case with tussock moths (*Lepidoptera*: *Lymantriidae*). Can J For Res, 2006, 36: 337–350
4 Besansky N J, Severson D W, Ferdig M T. DNA barcoding of parasites and invertebrate disease vectors: what you don't know can hurt you. Trends Parasitol, 2003, 19: 545–546
5 Bucklin A, Steinke D, Blanco-Bercial L. DNA barcoding of marine metazoa. Annu Rev Mar Sci, 2011, 3: 471–508
6 Eaton M J, Meyers G L, Kolokotronis S O, *et al*. Barcoding bushmeat: molecular identification of Central African and South American harvested vertebrates. Conserv Genet, 2010, 11: 1389–1404
7 Hebert P D, Cywinska A, Ball S L, *et al*. Biological identifications through DNA barcodes. Proc R Soc Lond, 2003, 270: 313–321
8 Lowenstein J H, Amato G, Kolokotronis S O. The real *maccoyii*: identifying tuna sushi with DNA barcodes—contrasting characteristic attributes and genetic distances. PLoS ONE, 2009, 4: e7866
9 Steinke D, Zemlak T S, Boutillier J A, *et al*. DNA barcoding of Pacific Canada's fishes. Mar Biol, 2009, 156: 2641–2647
10 Steinke D, Zemlak T S, Hebert P D. Barcoding nemo: DNA-based identifications for the ornamental fish trade. PLoS ONE, 2009, 4: e6300
11 Stoeckle M Y, Hebert P D. Barcode of life. Sci Am, 2008, 299: 88
12 Valentini A, Pompanon F, Taberlet P. DNA barcoding for ecologists. Trends Ecol Evol, 2009, 24: 110–117
13 Boehme P, Amendt J, Disney R H, *et al*. Molecular identification of carrion-breeding scuttle flies (Diptera: Phoridae) using COI barcodes. Int J Legal Med, 2010, 124: 577–581
14 Feng Y W, Li Q, Kong L F, *et al*. DNA barcoding and phylogenetic analysis of Pectinidae (Mollusca: Bivalvia) based on mitochondrial COI and 16S rRNA genes. Mol Biol Rep, 2011, 38: 291–299
15 Francis C M, Borisenko A V, Ivanova N V, *et al*. The role of DNA barcodes in understanding and conservation of mammal diversity in Southeast Asia. PLoS ONE, 2010, 5: e12575
16 Hebert P D, Ratnasingham S, de Waard J R. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc R Soc Lond B (Suppl), 2003, 270: S96–S99
17 Hebert P D, deWaard J R, Landry J F. DNA barcodes for 1/1000 of the animal kingdom. Biol Lett, 2009, 6: 359–362
18 Hebert P D, Stoeckle M Y, Zemlak T S, *et al*. Identification of birds through DNA barcodes. PLoS Biol, 2004, 2: e312
19 Hollingsworth P M, Forrest L L, Spouge J L, *et al*. A DNA barcode for land plants. Proc Natl Acad Sci USA, 2009, 106: 12794–12797
20 Zuo Y, Chen Z, Kondo K, *et al*. DNA barcoding of panax species. Planta Med, 2011, 77: 182–187
21 Chen S, Yao H, Han J, *et al*. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS ONE, 2010, 5: e8613
22 Luo K, Chen S, Chen K, *et al*. Assessment of candidate plant DNA barcodes using the Rutaceae family. Sci China Life Sci, 2010, 53: 701–708
23 Cräutlein M, Korpelainen H, Pietiläinen M, *et al*. DNA barcoding: a tool for improved taxon identification and detection of species diversity. Biodivers Conserv, 2011, 20: 373–389
24 Schwarz P, Bretagne S, Gantier J C, *et al*. Molecular identification of zygomycetes from culture and experimentally infected tissues. J Clin Microbiol, 2006, 44: 340–349
25 Druzhinina I S, Kopchinskiy A G, Komoń M, *et al*. An oligonucleotide barcode for species identification in *Trichoderma* and *Hypocrea*.

Fungal Genet Biol, 2005, 42: 813–828
26 Frøslev T G, Jeppesen T S, Læssøe T, *et al*. Molecular phylogenetics and delimitation of species in *Cortinarius* section *Calochroi* (*Basidiomycota*, *Agaricales*) in Europe. Mol Phylogenet Evol, 2007, 44: 217–227
27 Feau N, Vialle A, Allaire M, *et al*. Fungal pathogen (mis-) identifications: A case study with DNA barcodes on Melampsora rusts of aspen and white poplar. Mycol Res, 2009, 113: 713–724
28 Roe A D, Rice A V, Bromilow S E, *et al*. Multilocus species identification and fungal DNA barcoding: insights from blue stain fungal symbionts of the mountain pine beetle. Mol Ecol Resour, 2010, 10: 946–959
29 Zhao P, Luo J, Zhuang W Y. Practice towards DNA barcoding of the nectriaceous fungi. Fungal Diversity, 2011, 46: 183–191
30 Geiser D M, Jiménez-Gasco M, Kang S, *et al*. FUSARIUM-ID v.1.0: A DNA sequence database for identifying *Fusarium*. Eur J Plant Pathol, 2004, 110: 473–479
31 O'Donnell K, Cigelnik E. Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. Mol Phylogenet Evol, 1997, 7: 103–116
32 Seifert K A. Progress towards DNA barcoding of fungi. Mol Ecol Resour, 2009, 9: 83–89
33 Goffeau A, Barrell B G, Bussey H, *et al*. Life with 6000 genes. Science, 1996, 274: 563–567
34 Ma L J, Fedorovab N D. A practical guide to fungal genome projects: strategy, technology, cost and completion. Mycology, 2010, 1: 9–24
35 Cornell M J, Alam I, Soanes D M, *et al*. Comparative genome analysis across a kingdom of eukaryotic organisms: specialization and diversification in the fungi. Genome Res, 2007, 17: 1809–1822
36 Fitzpatrick D A, Logue M E, Stajich J E, *et al*. A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. BMC Evol Biol, 2006, 6: 99
37 Moran G P, Coleman D C, Sullivan D J. Comparative genomics and the evolution of pathogenicity in human pathogenic fungi. Eukaryot Cell, 2011, 10: 34–42
38 Soanes D M, Alam I, Cornell M, *et al*. Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis. PLoS ONE, 2008, 3: e2300
39 Firon A, d'Enfert C. Identifying essential genes in fungal pathogens of humans. Trends Microbiol, 2002, 10: 456–462
40 Jiang B, Bussey H, Roemer T. Novel strategies in antifungal lead discovery. Curr Opin Microbiol, 2002, 5: 466–471
41 Parkinson T. The impact of genomics on anti-infectives drug discovery and development. Trends Microbiol, 2002, 10: S22–S26
42 Hsiang T, Baillie D L. Comparison of the yeast proteome to other fungal genomes to find core fungal genes. J Mol Evol, 2005, 60: 475–483
43 Liti G, Louis E J. Yeast genome evolution and comparative genomics. Ann Rev Microbiol, 2005, 59: 135–153
44 Robert V, Szöke S, Eberhardt U, *et al*. The quest for a general and reliable fungal DNA barcode. Open Appl Inform J, 2011, 5: 45–61
45 Lewis C T, Bilkhu S, Robert V, *et al*. Identification of fungal DNA barcode targets and PCR primers based on Pfam protein families and taxonomic hierarchy. Open Appl Inform J, 2011, 5: 30–44
46 Hardison R C. Comparative Genomics. PLoS Biol, 2003, 1: e58
47 Rossman A Y, Samuels G J, Rogerson C T, *et al*. Genera of *Bionectriaceae*, *Hypocreaceae* and *Nectriaceae* (*Hypocreales*, *Ascomycetes*). Stud Mycol, 1999, 42: 1–248
48 Zhuang W Y. Taxonomy and related studies on the nectrioid fungi from China. Chin Bull Life Sci, 2010, 22: 1083–1085
49 Altschul S F, Gish W, Miller W, *et al*. Basic local alignment search tool. J Mol Biol, 1990, 215: 403–410
50 Hsiang T, Goodwin P H. Distinguishing plant and fungal sequences in ESTs from infected plant tissues. J Microbiol Methods, 2003, 54: 339–351
51 Thompson J D, Higgins D G, Gibson T J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res, 1994, 22: 4673–4680
52 Lalitha S. Primer Premier 5. Biotech Softw Internet Rep, 2000, 1: 270–272
53 Compton T. Degenerate primers for DNA amplification. In: Innis M A, Gelfand D H, Sninsky J J, *et al*., eds. PCR Protocols: A Guide to Methods and Applications. New York: Academic Press, 1990. 39–45
54 Dieffenbach C W, Dveksler G S. PCR Primer: a Laboratory Manual. New York: Cold Spring Harbor Press, 1995. 15–41
55 Goller S P, Gorfer M, Mach R L, *et al*. Gene cloning using PCR. In: Bridge P D, Arora D K, Reddy C A, *et al*., eds. Applications of PCR in Mycology. Wallingford: CAB International, 1998. 21–45
56 Innis M A, Gelfand D H. Optimization of PCRs. In: Innis M A, Gelfand D H, Sninsky J J, *et al*., eds. PCR Protocols: A Guide to Methods and Applications. New York: Academic Press, 1990. 3–12
57 Saiki R K. Amplification of genomic DNA. In: Innis M A, Gelfand D H, Sninsky J J, *et al*., eds. PCR Protocols: A Guide to Methods and Applications. New York: Academic Press, 1990. 13–20
58 Sambrook J, Russell D W. Molecular cloning: a laboratory manual. 3rd ed. New York: Cold Spring Harbor Press, 2001. 1.31–1.38
59 Wang L, Zhuang W Y. Designing primer sets for amplification of partial calmodulin genes from penicillia. Mycosystema, 2004, 23: 466–473
60 Hall T A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser, 1999, 41: 95–98
61 Slabbinck B, Dawyndt P, Martens M, *et al*. TaxonGap: a visualisation tool for intra- and inter-species variation among individual biomarkers. Bioinformatics, 2008, 24: 866–867
62 Tamura K, Dudley J, Nei M, *et al*. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol, 2007, 24: 1596–1599
63 Meyer C P, Paulay G. DNA barcoding: error rates based on comprehensive sampling. PLoS Biol, 2005, 3: e422
64 Hsiang T, Baillie D L. Recent progress, developments, and issues in comparative fungal genomics. Can J Plant Pathol, 2004, 26: 19–30
65 Pertsemlidis A, Fondon J W. Having a BLAST with bioinformatics (and avoiding BLASTphemy). Genome Biol, 2002, 2: 1–10
66 Zeng Z Q, Zhao F Y, Hsiang T, *et al*. Comparison of genomes between *Aspergillus nidulans* and 30 filamentous ascomycetes. Yi Chuan, 2010, 32: 1195–1202
67 Pearson W R. Empirical statistical estimates for sequence similarity searches. J Mol Evol, 1998, 276: 71–84
68 Keon J, Bailey A, Hargreaves J. A group of expressed cDNA sequences from the wheat fungal leaf blotch pathogen, *Mycosphaerella graminicola* (*Septoria tritici*). Fungal Genet Biol, 2000, 29: 118–133
69 Kruger W M, Pritsch C, Chao S, *et al*. Functional and comparative bioinformatic analysis of expressed genes from wheat spikes infected with *Fusarium graminearum*. Mol Plant Microbe Interact, 2002, 15: 445–455
70 Thomas S W, Glaring M A, Rasmussen S W, *et al*. Transcript profiling in the barley mildew pathogen *Blumeria graminis* by serial analysis of gene expression (SAGE). Mol Plant Microbe Interact, 2002, 15: 847–856
71 Thomas S W, Rasmussen S W, Glaring M A, *et al*. Gene identification in the obligate fungal pathogen *Blumeria graminis* by expressed sequence tag analysis. Fungal Genet Biol, 2001, 33: 195–211
72 Schubert K, Groenewald J Z, Braun U, *et al*. Biodiversity in the *Cladosporium herbarum* complex (Davidiellaceae, Capnodiales), with standardisation of methods for *Cladosporium* taxonomy and diagnostics. Stud Mycol, 2007, 58: 105–156
73 Skouboe P, Frisvad J C, Taylor J W, *et al*. Phylogenetic analysis of nucleotide sequences from the ITS region of terverticillate *Penicillium* species. Mycol Res, 1999, 103: 873–881
74 Liu Y J, Whelen S, Hall B D. Phylogenetic relationships among Ascomycetes: evidence from an RNA polymerase II subunit. Mol Biol Evol, 1999, 16: 1799–1808
75 Taberlet P, Coissac E, Pompanon F, *et al*. Power and limitations of the chloroplast *trn*L (UAA) intron for plant DNA barcoding. Nucleic Acids Res, 2007, 35: e14
76 Myers K K, Fonzi W A, Sypherd P S. Isolation and sequence analysis

of the gene for translation elongation factor 3 from *Candida albicans*. Nucleic Acids Res, 1992, 20: 1705–1710
77 Qin S L, Moldave K, McLaughlin C S. Isolation of the yeast gene encoding elongation factor 3 for protein synthesis. J Biol Chem, 1987, 262: 7802–7807
78 Ypma-Wong M F, Fonzi W A, Sypherd P S. Fungus-specific translation elongation factor 3 gene present in *Pneumocystis carinii*. Infect Immun, 1992, 60: 4140–4145
79 Fröhlich K U, Fries H W, Rüdiger M, *et al*. Yeast cell cycle protein CDC48p shows full-length homology to the mammalian protein VCP and is a member of a protein family involved in secretion, peroxisome formation, and gene expression. J Cell Biol, 1991, 114: 443–453
80 Imamura S, Ojima N, Yamashita M. Cold-inducible expression of the cell division cycle gene CDC48 and its promotion of cell proliferation during cold acclimation in zebrafish cells. FEBS Lett, 2003, 549: 14–20
81 Latterich M, Fröhlich K U, Schekman R. Membrane fusion and the cell cycle: Cdc48p participates in the fusion of ER membranes. Cell, 1995, 82: 885–893
82 Chen B, Piel W H, Gui L, *et al*. The HSP90 family of genes in the human genome: insights into their divergence and evolution. Genomics, 2005, 86: 627–637
83 Farrelly F W, Finkelstein D B. Complete sequence of the heat shock-inducible HSP90 gene of *Saccharomyces cerevisiae*. J Biol Chem, 1984, 259: 5745–5751
84 Nageshan R K, Roy N, Hehl A B, *et al*. Post-transcriptional repair of a split heat shock protein 90 gene by mRNA trans-splicing. J Biol Chem, 2011, 286: 7116–7122
85 Gupta R S. Phylogenetic analysis of the 90 kD heat shock family of protein sequences and an examination of the relationship among animals, plants, and fungi species. Mol Biol Evol, 1995, 12: 1063–1073
86 Gupta R S. Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. Mol Biol Rev, 1998, 62: 1435–1491
87 Stechmann A, Cavalier-Smith T. Evolutionary origins of Hsp90 chaperones and a deep paralogy in their bacterial ancestors. J Eukaryot Microbiol, 2004, 51: 364–373
88 Arends H, Sebald W. Nucleotide sequence of the cloned mRNA and gene of the ADP/ATP carrier from *Neurospora crassa*. EMBO J, 1984, 3: 377–382
89 Zimmerman R, Paluch U, Sprinzl M, *et al*. Cell-free synthesis of the mitochondrial ADP/ATP carrier protein of *Neurospora crassa*. Eur J Biochem, 1979, 99: 247–252