Conserved PCR Primer Set Designing for Closely-Related Species to Complete Mitochondrial Genome Sequencing Using a Sliding Window-Based PSO Algorithm

Cheng-Hong Yang1,2, Hsueh-Wei Chang3,4,5,6*, Chang-Hsuan Ho1, Yii-Cheng Chou7, Li-Yeh Chuang8*

1 Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan, 2 Department of Network Systems, Toko University, Chiai, Taiwan, 3 Department of Biomedical Science and Environmental Biology, Kaohsiung Medical University, Kaohsiung, Taiwan, 4 Graduate Institute of Natural Products, Kaohsiung Medical University, Kaohsiung, Taiwan, 5 Center of Excellence for Environmental Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan, 6 Cancer Center, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan, 7 Department of Medical Laboratory Science and Biotechnology, Chung Hwa University of Medical Technology, Tainan, Taiwan, 8 Department of Chemical Engineering & Institute of Biotechnology and Chemical Engineering, I-Shou University, Kaohsiung, Taiwan

Abstract

**Background:** Complete mitochondrial (mt) genome sequencing is becoming increasingly common for phylogenetic reconstruction and as a model for genome evolution. For long template sequencing, i.e., like the entire mtDNA, it is essential to design primers for Polymerase Chain Reaction (PCR) amplicons which are partly overlapping each other. The presented chromosome walking strategy provides the overlapping design to solve the problem for unreliable sequencing data at the 5’ end and provides the effective sequencing. However, current algorithms and tools are mostly focused on the primer design for a local region in the genomic sequence. Accordingly, it is still challenging to provide the primer sets for the entire mtDNA.

**Methodology/Principal Findings:** The purpose of this study is to develop an integrated primer design algorithm for entire mt genome in general, and for the common primer sets for closely-related species in particular. We introduce ClustalW to generate the multiple sequence alignment needed to find the conserved sequences in closely-related species. These conserved sequences are suitable for designing the common primers for the entire mtDNA. Using a heuristic algorithm particle swarm optimization (PSO), all the designed primers were computationally validated to fit the common primer design constraints, such as the melting temperature, primer length and GC content, PCR product length, secondary structure, specificity, and terminal limitation. The overlap requirement for PCR amplicons in the entire mtDNA is satisfied by defining the overlapping region with the sliding window technology. Finally, primer sets were designed within the overlapping region. The primer sets for the entire mtDNA sequences were successfully demonstrated in the example of two closely-related fish species. The pseudo code for the design algorithm is provided.

**Conclusions/Significance:** In conclusion, it can be said that our proposed sliding window-based PSO algorithm provides the necessary primer sets for the entire mt genome amplification and sequencing.

Citation: Yang C-H, Chang H-W, Ho C-H, Chou Y-C, Chuang L-Y (2011) Conserved PCR Primer Set Designing for Closely-Related Species to Complete Mitochondrial Genome Sequencing Using a Sliding Window-Based PSO Algorithm. PLoS ONE 6(3): e17729. doi:10.1371/journal.pone.0017729

Editor: Anita Kloss-Brandstätter, Innsbruck Medical University, Austria

Received September 30, 2010; Accepted February 12, 2011; Published March 18, 2011

Copyright: © 2011 Yang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work is partly supported by the National Science Council in Taiwan under grants NSC97-2311-B-037-003-MY3, NSC97-2622-E-151-008-CC2, NSC96-2221-E-124-050-MY3, NSC96-2311-B-037-002, NSC96-2622-E124-004-CC3, the funds DOH99-TD-C-111-002 and KMU-EM-99-1.4. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: changhw@kmu.edu.tw (H-WC); chuang@isu.edu.tw (L-YC)

Introduction

Mitochondrial DNA (mtDNA) is very popular for studying evolutionary relationships [1] because of its maternal inheritance and very high mutation rate, and lack of recombination [2,3,4,5]. Although establishing the phylogenetic tree is an important tool in understanding the evolutionary relationships among species [1], mitochondrial genomes depend solely on a small part of the entire mtDNA, such as *cytochrome b* (*cyt b*) [6,7], *cytochrome oxidase subunit I* (COI) [8,9], and others. These studies tend to underestimate the contribution of the variation of the entire mitochondrial genome in evolutionary processes.

For example, different parts in mtDNA sequences may show different mutation rates [10]. High homology often occurs in protein-coding genes and high variability appears in non-coding sequence segments [11,12]. Moreover, the mitochondrial protein-coding genes and the D-loop region evolve faster than 12S and 16S rRNA genes [13]. Hence, it is essential to address the phylogenetic relationship among species inferred from the different parts of mtDNA if the whole mt genomic sequences are available.

Although some mt genomic sequencing studies have been published [14,15,16,17,18,19,20], mitochondrial genome sequencing for most species is still incomplete as shown in the GenBank due to the technical problems related to the primer design. This is...
especially true for closely-related species. Previously, development of the conserved primers for rapid sequencing of the complete mitochondrial genome for several species, i.e., three species of bears, has been proposed [21]. However, the conserved primers were commonly designed by manual inspection of the prealigned mtDNA sequences, especially for the whole mitochondrial genome. To perform mitochondrial genome sequencing for many species without computation is still a challenge.

To solve these problems, we developed a heuristic approach, particle swarm optimization (PSO), coupled with the sliding window mechanism to design the most suitable primer sets for amplification of several similar mtDNA sequences after multiple sequence alignments from several closely-related species. PSO and the sliding window technique constitute an optimization technique and a randomized search respectively, and derive their working principles from the social behavior of organisms. Several important criteria in primer design, including the melting temperature, the PCR product size, the secondary structure, and the uniqueness of each designed primer [22] were considered in this study. Because the unreliable sequencing data at the start end of the first 30 to 40 nucleotides (nts), our proposed algorithm was designed to avoid these problems. Our strategy is to provide the primer sets that generate the PCR amplicons for each neighboring region with partial overlap. Accordingly, the unreliable sequencing data at the start end in one PCR amplicon can be compensated for by the 3’ end sequencing data from its upstream PCR amplicon.

Finally, we selected two sets of entire mt genomic sequences from two closely-related fish species to successfully demonstrate that our proposed algorithm was able to effectively identify the common primer sets for amplifying the entire mt genomic sequence.

Materials and Methods

Alignment of multiple sequences

Sequence alignment is the first step when designing a set of common PCR primer pairs of homologous sequences, which may be derived from closely-related species. A well-known multiple sequence alignment tool, ClustalW [23], was employed in this study. In the example of sequence alignment from three homologous sequences (Figure 1), the length of the match regions with larger than the constraint of the shortest primer length are regarded as the viable regions for primer design. After multiple sequences are aligned, numerous viable regions may be found and conserved. After the sliding window process is performed, the forward and reverse primers within the overlapping and conserved regions can easily be designed to amplify the region between them.

Primer design using particle swarm optimization

Primer design is of crucial importance for PCR experiments. The quality of primers always influences whether a PCR experiment is successful or not. To obtain high quality primers, many primer design constraints must be satisfied. A heuristic algorithm particle swarm optimization (PSO) is employed. Particle swarm optimization is a population-based stochastic optimization technique, which was developed by Kennedy and Eberhart in 1995 [24]. PSO simulates the social behavior of organisms, such as birds in a flock and fish in a school. This behavior can be described as an automatically and iteratively updated system. In PSO, each single candidate solution can be considered “an individual bird of the flock”, that is, a particle in the search space. Each particle makes use of its own memory and knowledge gained by the swarm as a whole to find the best (i.e., optimum) solution. All of the particles have fitness values, which are evaluated by a fitness function so that they can be optimized. During movement, each particle adjusts its position by changing its velocity according to its own experience and according to the experience of a neighboring particle, thus making use of the best position encountered by itself and its neighbor. Particles move through the problem space by following a current of optimum particles. The process is then reiterated a fixed number of times or until a predetermined minimum error is achieved [24]. The computational flowchart of the PSO is provided (Figure 2). Details of the PSO algorithm are discussed in the following sections.

Encoding schemes

As stated above, we employed PSO to design suitable primer sets. In the PSO, each particle is designed in a format that enabled us to express a particular primer pair. The encoding method used was the following:

\[ P_i = \{F_i, F_t, R_t, P_i\}, i = 1, 2, ..., n \]  

where \( n \) represents the size of the population, \( F \) represents the start index of the forward primer, \( F_t \) represents the length of forward primer, \( R_t \) represents the length of the reverse primer, and \( P_i \) represents the length of the PCR product. In PSO, a particle thus described represents a primer pair in a specific window. A particle represented by \( P_i = \{2, 3, 4, 10\} \) as shown in Figure 3 as an example. Based on the template sequence, the \( i \)-th particle representing the PCR product is “CTTAGCGAAT” in which the forward and reverse primer are “CITT” and “ATTC” (with complementary to reverse primer of GAAT), respectively.

Population initialization

To initialize a population, all of the particles are randomly generated and each particle is given a velocity \((v)\) within \(0 \rightarrow 1\). Initially, \( F \) is randomly generated in the template sequence length of the window, and \( F_t \) and \( R_t \) are generated within the constraints of the primer length. \( P_i \) is randomly generated within the constraints of the PCR length. The velocity of each dimension of each particle is randomly generated.

![Figure 1. Identification of suitable primer design regions from a ClustalW generated sequence alignment (The parameter setting for the minimum primer length was 18).](https://doi.org/10.1371/journal.pone.0017729.g001)
Fitness function

A successful PCR experiment depends on the quality and specificity of the designed primers. The optimal primer must satisfy many criteria [25]. In PSO, a fitness function is used to evaluate the fitness of each particle in order to check whether the primer pairs satisfy the design constraints. We combined many primer design constraint functions with weights into the fitness function. The constraints used are the followings:

1) melting
Conserved Primer Design for Whole mtDNA Sequencing

temperature, 2) primer length and GC content, 3) PCR product length, 4) secondary structure, 5) specificity, and 6) terminal limitation. These constraints are described in detail below:

1) Melting temperature. To perform a successful PCR experiment, the melting temperature (T_m) for each primer must be confined to a suitable range. In this study, the value of the melting temperature of the primer is denoted T_m(P_i), referring to the formula based on nearest neighbor thermodynamic theory by Freier et. al (Eq. (2)) [26,27] and adopted from the user manual of “NetPrimer” software from PREMIER Biosoft International (http://www.premierbiosoft.com/DATAFILES/NetPrimerManual.pdf). In Eq. (2), ΔH and ΔS indicate enthalpy and entropy for helix formation, respectively, which are both calculated as the absolute value.

\[
T_m(\text{primer}) = \Delta H / (\Delta S + R \times \ln (C/4)) + 16.6 \log ([K^+]/(1 + 0.7[K^+])) - 273.15
\] (2)

\[
\Delta H = \sum_{i=1}^{n} \Delta H(\text{interaction between base}_i \text{ and base}_{i+1})
\] (3)

which \( n \) is sequence selected length.

\[
\Delta S = \sum_{i=1}^{n} \Delta S(\text{interaction between base}_i \text{ and base}_{i+1})
\] (4)

which \( n \) is sequence selected length.

\[
[K^+] = [\text{Monovalent ion concentration}] + 4 \times \sqrt{\text{(Free}[\text{Mg}^{2+}] \text{ ion concentration} \times 1000)}
\] (5)

\[
Melt_{\text{tm}}(P_i) = \begin{cases} 
0, & \text{if } 54 \leq T_m(P_{iF}), T_m(P_{iR}) \leq 65 \\
1, & \text{otherwise}
\end{cases}
\] (6)

\[
\Delta Melt_{\text{tm}}(P_i) = \begin{cases} 
0, & \text{if abs}(T_m(P_{iF}) - T_m(P_{iR})) \leq 3 \\
1, & \text{otherwise}
\end{cases}
\] (7)

2) Primer length and GC content. In general, the primer length should be within 18 and 28 nts, and the differential length of a primer pair is restricted to less than 3 nts [29]. \( |P_{iF}| \) and \( |P_{iR}| \) represent the number of nucleotides of the forward and reverse primers, respectively. In Eq. (8) and (9), the Length\( (P_i) \) is used to check whether the length of a primer pair is within 18 to 28 nts, and \( \Delta \text{Length}(P_i) \) is used to check whether the length difference between the forward and reverse primers exceeds 3 nts or not.

\[
\text{Length}(P_i) = \begin{cases} 
0, & \text{if } 18 \leq |P_{iF}|, |P_{iR}| \leq 28 \\
1, & \text{otherwise}
\end{cases}
\] (8)

\[
\Delta \text{Length}(P_i) = \begin{cases} 
0, & \text{if abs}(|P_{iF}| - |P_{iR}|) \leq 3 \\
1, & \text{otherwise}
\end{cases}
\] (9)

The GC content check function \( \text{GC}\% (P_i) \) is denoted as Eq. (7)

\[
\text{GC}\% (P_i) = \begin{cases} 
0, & \text{if } 40\% \leq \text{GC}\% (P_{iF}), \text{GC}\% (P_{iR}) \leq 60\% \\
1, & \text{otherwise}
\end{cases}
\] (7)

where \( \text{GC}\% (\text{primer}) = \#G(\text{primer}) + \#C(\text{primer}) \nabla \text{length of sequence})\]

3) PCR product length. MtDNA is a double-stranded circular molecule. PCR amplicons using several primer pairs are surrounded by the entire mtDNA sequences without gap. It is also essential to maintain the coverage for assembly of the individual sequences from different amplicons into the complete mt genome. The length of the PCR product has to be considered. In Eq. (10), \( |P_{iF}| \) is the PCR product length. Using forward and reverse primers to perform the two-directional sequencing are helpful for double checking the sequence for 800–1100 nts in reliability. When the PCR product length setting of 800–1100 nts is not
available, it is adjustable to reduce or extend the PCR length to a suitable range until it is reached.

\[
\text{productLength}(P_i, P_f) = \begin{cases} 
0, & \text{if } 800 \leq |P_i| \leq 1100 \\
1, & \text{otherwise} 
\end{cases} \tag{10}
\]

4) Secondary structure. In primer design, primers that bind to any sites on the sequence indiscriminately have to be avoided. Furthermore, primers that are self-complementary (self-dimer) or complement each other (cross-dimer) must also be avoided where the dimer was defined to possess over 5 base pairings and the hairpin length had to longer than 4 bps. These constraints are defined in Eq. (11) and Eq. (12):

\[
\text{Dimer}(P_i) = \begin{cases} 
0, & \text{if } P_{id} \text{ and } P_{ir} \text{ are not self-complementary or do not complement each other} \\
1, & \text{otherwise} 
\end{cases} \tag{11}
\]

\[
\text{Hairpin}(P_i) = \begin{cases} 
0, & \text{if primer doesn’t complement itself} \\
1, & \text{in the U form of } P_{id} \text{ and } P_{ir} \\
1, & \text{otherwise} 
\end{cases} \tag{12}
\]

5) Specificity. The specificity constraint is used to judge whether the sequence repeatedly occurs in primer or not in order to ensure the specificity of the primer. The PCR experiment might fail if the primer is not site-specific or appears more than once in the sequence.

\[
\text{Unipair}(P_i) = \begin{cases} 
0, & \text{if } P_{id} \text{ and } P_{ir} \text{ appear in the template sequence of the window once} \\
1, & \text{otherwise} 
\end{cases} \tag{13}
\]

6) Terminal limitation. \(\text{GC_clamp}(P_i)\) is used to judge whether the nucleotides located in the 3’ end of primer are G or C:

\[
\text{GC_clamp}(P_i) = \begin{cases} 
0, & \text{if 3’ end of } P_{id} \text{ and } P_{ir} \text{ is G or C} \\
1, & \text{otherwise} 
\end{cases} \tag{14}
\]

7) End match. \(\text{Endmatch}(P_i)\) is used to judge whether the three nucleotides located in the 3’ end of primer are base pairings:

\[
\text{Endmatch}(P_i) = \begin{cases} 
0, & \text{if 3’ end of three nucleotides are base pairings} \\
1, & \text{otherwise} 
\end{cases} \tag{15}
\]

8) Terminus GC number limitation. \(\text{GC_Terminus Limitation}(P_i)\) is used to judge whether the five nucleotides located in the 3’ and 5’ termini of primer occur not over 3G or 3C:

\[
\text{GC_Terminus Limitation}(P_i) = \begin{cases} 
0, & \text{if 3’ and 5’ termini of five nucleotides occur} \\
1, & \text{otherwise} 
\end{cases} \tag{16}
\]

Based on the several primer design constraints described above, the fitness of each particle is evaluated by the fitness function. A low fitness value means that the particle fits more constraints. The default fitness function can be written as:

\[
\text{Fitness}(P_i) = 10 \times \text{productLength}(P_i, P_f) + \text{GC_clamp}(P_i) + \text{Dimer}(P_i) + \text{Hairpin}(P_i) + 20 \times (\text{Length}(P_i)) + \Delta \text{Length}(P_i) + \Delta \text{Melt_tm}(P_i) + \Delta \text{Melt_tm}(P_f) + 
\frac{\text{GC}\% (P_i) + \text{GC}\% (P_f)}{2} + 50 \times (\text{Unipair}(P_i) + \text{Endmatch}(P_i))
\]

All the constraints of multiplex PCR primer design are combined into an objective function with weights. Every constraint weight setting is based on our previous researches [30,31].

Particle update

One of the characteristics of PSO is that each particle has a memory of its own best experience. At every iteration the particle’s trajectory is updated by two “best” values, called \(pbest\) and \(gbest\). Each particle keeps track of its coordinates in the problem space, which are associated with the best solution (fitness) the particle has achieved so far. This fitness value is stored, and represents the position called \(pbest\). When a particle takes the whole population as its topological neighborhood, the best value is a global optimum value called \(gbest\).

In this study, the adaptive functional values were based on the particle features representing the feature dimension; this data was evaluated by several constraints to obtain a fitness value. Once the adaptive values \(pbest\) and \(gbest\) are obtained, the features of the \(pbest\) and \(gbest\) particles can be tracked with regard to their position and velocity. Each particle is updated according to the following equations.

\[
\begin{align*}
\text{v}^\text{new} & = w \times \text{v}^\text{old} + c_1 \times \text{rand}_1 \times (\text{pbest}_\text{id} - \text{x}^\text{old}) + c_2 \times \text{rand}_2 \times (\text{gbest}_\text{id} - \text{x}^\text{old}) \tag{18} \\
\text{if } \text{v}^\text{new} \neq (\text{V}_{\text{min}}, \text{V}_{\max}) & \text{ then } \text{v}^\text{new} = \max(\min(\text{V}_{\max}, \text{v}^\text{new}), \text{V}_{\min}) \tag{19} \\
\text{x}^\text{new} & = \text{x}^\text{old} + \text{v}^\text{new} \tag{20}
\end{align*}
\]

In these equations, \(w\) is the inertia weight, \(c_1\) and \(c_2\) are acceleration (learning) factors, and \(\text{rand}_1\) and \(\text{rand}_2\) are random numbers between 0 and 1. Velocities \(\text{v}^\text{old}\) and \(\text{v}^\text{new}\) are those of the updated particle and the particle before being updated, respectively, \(\text{x}^\text{old}\) is the original particle position (solution), and \(\text{x}^\text{new}\) is the updated particle position (solution).

In Eq. (19), particle velocities of each dimension are tried to a maximum velocity \(\text{V}_{\max}\). If the sum of the accelerations causes the velocity of that dimension to exceed \(\text{V}_{\max}\), the velocity of that dimension is limited to \(\text{V}_{\max}\). \(\text{V}_{\max}\) and \(\text{V}_{\min}\) are user-specified parameters.

Primer design based on the sliding window technique

The sliding window method is simple and quick technique and has been widely applied in several computation fields, such as 16S ribosomal DNA amplicons in metagenomic studies [32], primer set selection in multiple PCR experiments [33], real-time primer
design for DNA chips [34], and accurate location of eukaryotic protein coding regions [35]. The basic concept defines a window which is sliding across the template sequence as a specified shift value. In multiple primer design problems, the window is slid using each previous primer pair at a time along the template sequence. Subsequently, a set of primer pairs can be designed for the coverage of the entire mt genome. A detailed diagram for the sliding window method is provided (Figure 4). First, the window starts from the first nucleotide of the conserved sequences and the PSO algorithm applies it to the designed primer pair within. Then the window shifts based on the previous (first) reverse primer and the next (second) forward primer is designed from the upstream region of the first reverse primer. Each shift provides enough length for fitting PCR product coverage constraints (ranging from 90 to 200 nts of the upstream of the reverse primer). When primers cannot be designed within this constrain, the overlapping length for each adjacent PCR product is extended automatically until the primers are suitable. This procedure continues until the entire template sequence is covered by all primer pair sets. The remaining primers are designed in the same way. The pseudo-code of the sliding window method based PSO primer design is provided (Figure 5). The brief JAVA 6.0-based software solution that implements our proposed algorithm was provided in the Supporting Information (Documentation S1 and Software S1).

Results

In this research project, we applied a high-performance PSO algorithm to design fit primer pairs, and used the sliding window technique to cover the entire mt genomic sequences. The test data sets and the detailed experiment result are described below.

Parameter settings

The termination condition of the PSO in this study is reached at a pre-specified number of iterations (in our case the number of iterations was 50). Parameters used here were a population size of 20, rand<sub>1</sub> and rand<sub>2</sub> were random numbers between (0, 1), and c<sub>1</sub> and c<sub>2</sub> were set to c<sub>1</sub> = c<sub>2</sub> = 2. The inertia weight w was 0.8. The velocity constraints were set to V<sub>max</sub> = 6 and V<sub>min</sub> = −6.

Test data

Entire mitochondrial genomic DNA sequences of two closely-related fish species, i.e., Scarus forsteni (FJ619271.1) and Scarus rubroviolaceus (FJ227899.1) are used as an example for designing the conserved PCR primer sets for closely-related species using a PSO algorithm. These mt genomic sequences were downloaded from NCBI GenBank.

Dry experiment

After multiple sequences were aligned, we implemented the PSO algorithm and the sliding window technique to examine the performance on biological test data, as described above. A set of primers for amplifying entire the mt genome which was approximately 16 kb long, was designed using the proposed algorithm.

As shown in Table 1, these primers had a similar length, GC content and annealing temperature and the constraints fit the general primer design restrictions. The forward 5'-ATTAGCC-CAGGACACCTAGGC-3' and the reverse 5'-GATTGGCACCGGTATTTTCTC-3' of the primer pair No. 1, the length (nts), GC% and Tm (°C) were 22, 43.5, 40.9 and 56.2 and 57.3, respectively. All primer sets listed in Table 1 obey the common primer constraints, i.e. hairpin, dimer and specificity. The PCR amplification was easy to complete. The adjacent PCR product overlap constraint in the proposed algorithm was set to 90−200 nts as mentioned in the materials and methods section. The overlapping lengths are 104 ( = 1058 vs. 956) and 165 nts ( = 1956 vs. 1816) for primer sets 1 and 2 and primer sets 2 and 3, respectively. Except for the primer sets 4/5, 6/7, 7/8, 8/9, 14/15, 19/20, 21/22 and 22/1, the overlapping lengths for the other primers were all in the range of 90 to 200 nts. When suitable primer pairs in the pre-set overlap length range were unavailable, the overlapping lengths of some adjacent PCR products were raised automatically. The overlapping lengths for primer sets 4/5, 6/7, 7/8, 8/9, 14/15, 19/20, 21/22 and 22/1 for example, were 207, 256, 310, 247, 362, 301, 243 and 398 nts, respectively.

In order to amplify the entire mt genome, it is essential to design a group of primer pairs that are partly overlapping each other. Based on the sliding window technique, the proposed algorithm can easily design primers that allow PCR amplification for 22 adjacent PCR products that overlap each other. As shown in Figure 6, the proposed algorithm yields the entire sets of forward and reverse primers. The expected length for these PCR products ranged from 800 to 1100 nts, and these primer sets are arranged neatly and overlap with the adjacent PCR products. All designed forward and reverse primers of each PCR product are designed within the conserved regions of the multiple sequence alignment.

Wet experiment

The first ten primer sets for mitochondrial genome of Scarus forsteni provided in Table 1 were examined by PCR amplification. DNA samples (200 ng) were added to the PCR reaction mixture (10 μl) containing 1 μl of 10× PCR buffer, 0.3 μl of 50 mM MgCl<sub>2</sub>, 0.2 μl of 10 mM dNTP each, 0.6 μl of DMSO, 0.14 μl of 5 U Platinum Taq enzyme (Invitogen corp.), 0.12 μl of 350 μg/ml primer mix [1:1], and 7.64 μl of DNA in water. The touch-down PCR program [36] was applied to all the test primer sets (listed in Table 1, sets 1 to 10) and it was slightly modified as follows: 94°C (3 min); 3 cycles of 94°C (15 s), 63°C (15 s), 70°C (30 s); 3 cycles of 94°C (15 s), 60°C (15 s), 70°C (30 s); 3 cycles of 94°C (15 s), 57°C (15 s), 70°C (30 s); 49 cycles of 94°C for (15 s), 54°C (15 s), 70°C (30 s). To avoid the very high risk of sample mixup and the resulting artificial recombination when using e.g. many different primer pairs (see Table 1), however, parallel amplification of all targets per specimen would be highly recommended. As shown in Figure 7, the designed primer sets were successfully amplified by PCR.

Discussion

The public resources available for mitochondrial bioinformatics have been reviewed [37]. Among them, only V-MitoSNP [36] provides primer design for mitochondrial genome sequences.
However, V-MitoSNP is still limited to mtSNPs. Currently, several primer design algorithms are used, e.g. genetic algorithm (GA) [29,30], heuristic algorithm [25], memetic algorithm (MA) [30], PSO algorithm [31], and consecutive multiple discovery (CMD) algorithm [39]. However, these algorithms were developed for primer design without considering the situations in multi-sequence alignment, and for the circular sequences like the mt genome. In contrast, some sequence alignment tools, such as CLUSTALW [40], TBA [41], MAVID [42], MLAGAN [43], Pecan [44], SeqSNPing [45], AQUA [46], and Cgaln [47], were developed without a primer design function. Accordingly, the development of integrated algorithms for designing feasible primer sets for entire mt genome sequences coupled with the sequence alignment are still challenging.

Recently, GeneFisher2 [48] was developed to design a number of primer pairs for consensus sequence after multi-sequence alignment. However, like most primer design tools, GeneFisher2 only focuses on designing a primer pair or a number of primer pairs on a limited region and does not design a set of primer pairs for amplifying an entire sequence at once. Primer design for

Table 1. A set of primers for amplifying entire circular mtDNA.

| No. | Forward Primer                                      | Reverse Primer                                      | Product Length |
|-----|-----------------------------------------------------|-----------------------------------------------------|----------------|
| 1   | ATTAGGCAATGACACTTAGCC                               | GCATTTGCACGATTTTCTT                                  | 1058           |
| 2   | GAGTTGGCCATCACCCAG                                  | GCCAGGTGGTTGGCTTCTT                                  | 1956           |
| 3   | TGTATTGCAAGATAGCC                                   | GATTTGCAAGCTTAGG                                    | 1816           |
| 4   | TTGACATGAAGATAGCC                                   | CACACATGGAAGACTTCC                                   | 2531           |
| 5   | TACCATGGATCAGGAG                                    | GCAGTTGGAAGACTTCC                                   | 5133           |
| 6   | CATGATGACAGGAG                                      | GCTAGTTGGAAGACTTCC                                   | 4995           |
| 7   | CATATGATGAGAG                                      | TCTGATGCAAGACTTCC                                   | 5555           |
| 8   | GCTGACATGACGAG                                      | GCAGTTGGAAGACTTCC                                   | 6152           |
| 9   | GAATGATGAGAG                                      | TTTTGGGTTTTTTC                                      | 7118           |
| 10  | CTAATGATGACGAG                                      | GCAGTTGGAAGACTTCC                                   | 8006           |
| 11  | CATCAGATTGAGAG                                      | GCAGTTGGAAGACTTCC                                   | 8731           |
| 12  | ATTTGAGGAGAG                                      | GCAGTTGGAAGACTTCC                                   | 9569           |
| 13  | GAGTTGAGGAGAG                                      | GCAGTTGGAAGACTTCC                                   | 10418          |
| 14  | CATCAGATTGAGAG                                      | GCAGTTGGAAGACTTCC                                   | 11011          |
| 15  | AGAGTTGAGGAGAG                                     | GCAGTTGGAAGACTTCC                                   | 11857          |
| 16  | CATCAGATTGAGAG                                      | GCAGTTGGAAGACTTCC                                   | 12815          |
| 17  | TCCATAGGAGAGAG                                     | GCAGTTGGAAGACTTCC                                   | 13497          |
| 18  | AGAGTTGAGGAGAG                                     | GCAGTTGGAAGACTTCC                                   | 14349          |
| 19  | AAATGATGAGGAG                                      | GCAGTTGGAAGACTTCC                                   | 14903          |
| 20  | GCTTATGAGGAGAG                                     | GCAGTTGGAAGACTTCC                                   | 15644          |
| 21  | ACTTATGAGGAGAG                                     | GCAGTTGGAAGACTTCC                                   | 16470          |

*The position of the first nucleotide in primer. The nucleotide “1” is the first nucleotide in the GenBank accession nos as described in Section materials and methods. doi:10.1371/journal.pone.0017729.t001
amplification and sequencing of an entire mt genome is currently under development.

PSO has been reported to progress faster than genetic algorithm (GA) with crossover, mutation, or both [49]. In our previous work for primer design [31], we found that the average accuracy of PSO was better than memetic algorithm (MA) [30] and genetic algorithm (GA) based on the Wallace formula [50] (100% vs. 98.33% and 74.93%) and the Bolton and McCarthy formula [51] (94.93% vs. 88.93% and 32.40%) for five hundred runs with PCR product length between 150~300 nts, 500~800 nts and 800~1000 nts. Therefore, we applied the PSO rather than the MA and GA to design the whole set of primers for the whole mt genomic sequencing.

In this study, we propose a novel strategy that introduces the sliding window technique coupled with a PSO algorithm. This algorithm provides a sequence alignment function for the entire circular genome and designs a set of primers that generates overlapping PCR amplification for the entire circular genome, e.g. the mt genome. Several primer design constraints for a successful PCR were considered in the proposed algorithm. For example, the melting temperature, the primer pair length, the GC content, the PCR product length, the secondary structure, and the specificity were all considered. The primer design results demonstrated that a set of primers that obeys these design constrain with suitable length and could be found for the PCR amplicons. Other circular genome sequences such as chloroplasts and bacterial chromosomes could be used with this strategy as well.

Some entire mt genome sequences are available for some species in GenBank, and hence their requirement for sequencing is reduced. However, the known mtDNA sequences could be applied to the entire mt genomic sequences of closely-related species based on our experience. For the purpose of sequencing species with an unknown mt genomic sequence, we suggest to collect the entire mt genomic sequences of the same family or order and perform our proposed algorithm. Subsequently, the designed primer sets derived from the conserved regions stand a high possibility of being used to perform successful PCR reactions for the entire mt genomic sequencing to the species with unknown mtDNA sequence.

Utilizing the heuristic algorithm and sliding window technique, we used several primer constraints to appraise the fitness values and, based on their respective significance, each constraint was given a corresponding weight. Through the design of a fitness function, our algorithm was able to design a complete set of primers. The strategy of this algorithm was designed to carefully...
select a set of common PCR primer pairs from the conserved regions by multiple nucleotide sequence alignment. The common PCR primer pairs designed by our algorithm could be applied to amplify conserved sequences of genes from the same or different organisms. It can be used to amplify conserved sequences of genes from other families. For example, the whole primer sets for amplifying the entire circular mtDNA for Chimpanzees (Pan troglodytes; accession no. NC 001643) [52] and Bonobo (Pan paniscus; accession no. GU189657) [53] are successfully mined using our proposed methodology (data not shown). In our study, two fish mitochondrial genome data sets were used to test the proposed method and ten of the whole primer sets were successfully amplified to prove the performance of this algorithm.

Taken together, our proposed primer design algorithm is an effective method to design primer pair sets for PCR amplification and sequencing for the multiple alignments of circular consensus sequences. The entire mt genome sequencing was easy to complete and could help biologists in identifying the phylogenetic relationship for closely-related species.

References

1. Moyle RG, Marks BD (2006) Phylogenetic relationships of the bulbuls (Aves: Pycnonotidae) based on mitochondrial and nuclear DNA sequence data. Mol Phylogenet Evol 40: 687–695.
2. Avise JC (1986) Mitochondrial DNA and the evolutionary genetics of higher animals. Philos Trans R Soc Lond B Biol Sci 312: 325–342.
3. May-Collado L, Agnarsson I (2006) Cytochrome b and Bayesian inference of evolutionary relationships in Phasianidae. Mitochondrial DNA 21: 5–7.
4. Chuang LY, Yang CH, Cheng YH, Gu DL, Chang PL, et al. (2006) V-primers for rapid sequencing of the complete mitochondrial DNA of mammals. BMC Evolutionary Biology 6: 8–9.
5. Nabholz B, Glemin S, Galtier N (2009) The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. BMC Evolutionary Biology 9: 54.
6. Webb DM, Moore WS (2005) A phylogenetic analysis of woodpeckers and their allies using 12S, 16S and COI nucleotide sequences (class Aves; order Piciformes). Mol Phylogenet Evol 36: 233–248.
7. Kerr KCR, Lijtmaer DA, Barreira AS, Hebert PDN, Tubaro PL (2009) Probing the relationship of rock carp Procypris rabaudi (Cypriniformes: Cyprinidae) and phylogenetic sequence variations of nuclear genes. Gene 429: 31–36.
8. Wu JS, Lee C, Wu CC, Shiau SH, Liu YC, Lin YM, et al. (2004) Primer design using genetic algorithm. Chemical and Biological Engineering 3: 18–23.
9. Simmler H, Singpiel H, Ma¨nner R (2004) Real-time primer design for DNA chips. Concurreny Computat: Pract Exper 16: 855–872.
10. Lee KF, Liu WL, Chen Y, Fang L, Zhang X, et al. (2008) Identification of complete mitochondrial genome of the tufted deer. Mitochondrial DNA 19: 411–417.
11. Dehle J, Strobeck C (2002) Conserved primers for rapid sequencing of the complete mitochondrial genome from carnivores, applied to three species of bears. Mol Biol Evol 19: 357–361.
12. Fernandes RJ, Skiena SB (2008) Microarray synthesis through multiple-use PCR primer design. Bioinformatics 18 Suppl 1: i218–i235.
13. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680.
14. NLR, He, I. (2004). The thermodynamics of DNA structural motifs. Annu Rev Biophys Biomol Struct 33: 413–440.
15. Breslauer KJ, Frank R, Blocker H, Marky LA (1986) Predicting DNA duplex stability from the base sequence. Proc Natl Acad Sci U S A 83: 3746–3750.
16. Wu JS, Lee C, Wu CC, Shiau SH, Lin YM, et al. (2004) Primer design using genetic algorithm. Bioinformatics 20: 1710–1717.
17. Yang CH, Cheng YH, Chuang LY, Chang HW (2009) Specific PCR product primer design using meristic algorithm. BioTechnol Prog 25: 743–753.
18. Yang CH, Cheng YH, Chang HW, Chuang LY (2010) Primer designer with specific PCR product using particle swarm optimization. International Journal of Chemical and Biological Engineering 3: 18–23.
19. Wang Y, Qian FY (2009) Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. PLoS ONE 4: e4701.
20. Lai WT, Yang CB, Shiau SH, Shih YL (2005) Primer selection using a novel multiple PCR primer design method for 16S rRNA genes. Nucleic Acids Res 33: 181–187.
41. Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, et al. (2004) Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res 14: 708–715.
42. Bray N, Pachter L (2004) MAVID: constrained ancestral alignment of multiple sequences. Genome Res 14: 693–699.
43. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, et al. (2003) LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res 13: 721–731.
44. Paten B, Herrero J, Beal K, Fitzgerald S, Birney E (2008) Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment paralogs. Genome Res 18: 1814–1828.
45. Chang HW, Chuang LY, Cheng YH, Ho CH, Wen CH, et al. (2009) SeqSNPIng: multiple-alignment tool for SNP discovery, SNP ID identification, and RFLP genotyping. OMICS 13: 253–260.
46. Muller J, Creevey CJ, Thompson JD, Arendt D, Bork P (2010) AQUA: automated quality improvement for multiple sequence alignments. Bioinformatics 26: 263–265.
47. Nakato R, Gotoh O (2010) Cigaln: fast and space-efficient whole-genome alignment. BMC Bioinformatics 11: 224.
48. Lamprecht AL, Margaria T, Steffen B, Sczyrba A, Hartmeier S, et al. (2008) GeneFisher-P: variations of GeneFisher as processes in Bio-jETI. BMC Bioinformatics 9 Suppl 4: S13.
49. Poli R, Kennedy J, Blackwell T (2007) Particle swarm optimization. An overview. Swarm Intell 1: 33–57.
50. Wallace RB, Shaffer J, Murphy RF, Bonner J, Hirose T, et al. (1979) Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch. Nucleic Acids Res 6: 3543–3557.
51. Bolton ET, Mc CB (1962) A general method for the isolation of RNA complementary to DNA. Proc Natl Acad Sci U S A 48: 1390–1397.
52. Stone AC, Battistuzzi FU, Kubatko LS, Perry GH, Jr., Trudeau E, et al. (2010) More reliable estimates of divergence times in Pan using complete mtDNA sequences and accounting for population structure. Philos Trans R Soc Lond B Biol Sci 365: 3277–3288.
53. Zsurka G, Kudina T, Peercy V, Hallmann K, Elger CE, et al. (2010) Distinct patterns of mitochondrial genome diversity in bonobos (Pan paniscus) and humans. BMC Evol Biol 10: 270.